FACTORS INFLUENCING THE APPLICATION OF COST-UTILITY ANALYSIS IN ASSESSING DISEASE MANAGEMENT AMONG ASIAN PATIENTS

ZHANG XUHAO

BSc. (Pharm), FU DAN UNIVERSITY

A THESIS SUBMITTED FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHARMACY

NATIONAL UNIVERSITY OF SINGAPORE

2008

Acknowledgements

Completing this thesis has been a very challenging but always inspiring journey. At the very beginning of this thesis, I would like to take this opportunity to express my most sincere gratitude and appreciation to all those who have given me kind support throughout this journey.

I am immensely grateful to my wonderful supervisors, Prof. Li Shu-Chuen and Prof. Julian Thumboo, for their tremendous dedication throughout my Ph D studies. They have offered me invaluable guidance and encouragement on how to be a capable and independent investigator, which has become and would remain as the most important treasure to me and my entire career. I also feel very indebted to them, as they have also sacrificed a lot of their personal time to thoroughly review and comment on various study protocols, manuscripts, and thesis, which enabled me to improve continuously. Special thanks also go to Prof and Mrs. Li for all their care and love to me, not only as supervisor but also just like my parents.

I am very grateful to all my collaborators from different institutions and organizations, Prof. Fong Kok-Yong, Drs Lo Ngai-Nung, Yeo Seng-Jin, Yang Kuang-Ying, Tan Hwee-Huan (Singapore General Hospital), Dr. Kevin Tan (Diabetic Society of Singapore), Dr. Wee Hwee Lin and Ms. Jin Jing (National University of Singapore) and Dr. Xie Feng (McMaster University, Canada). They have given me great insights, valuable comments and kind support throughout the collaboration.

Many thanks are also due to Assoc. Prof. Paul Ho and Asst Prof. Chui Wai Keung (my Ph D thesis advisory committee), Assoc Prof. Chan Sui Yung (Head of Department), and Assoc Prof Chan Lai Wah (Chairman of Pharmacy Graduate Program Committee), and all the other staff and fellow students in my department for their crucial support and invaluable advice throughout my Ph D studies. I also would

like to thank the Department of Pharmacy and International Society of Pharmacoeconomics and Outcomes Research for awarding me the research scholarship and travel grants.

As I joined Merck Sharp & Dhome at the end of the 3rd year, I really would like to appreciate the tremendous understanding and support from my managing director (Ms. Annie Chin), my ex-medical director (Dr. Lai Hung Jen), my manager (Mr. Lee Jet Tong) and all the other fellow colleagues.

Last but not least, my deepest gratitude goes to my parents (Mr. Zhang Ji Hua and Mdm Xiao Ming Qi), my husband (Mr. Leon Xu), my very cute "Shitzu" sister (Sissy) and all the other family and extended family members and friends for their endless love and unflagging support throughout my life.

Table of Contents

SUMMARY

Due to the aggregated impact of rising chronic disease prevalence and increasing demand for better quality of health care, disease management has been increasingly emphasized across Asia to achieve optimal health outcomes at affordable expenditures. In order to ensure that disease management programs are realizing their value-adding capacities to health care systems, it is essential to apply economic evaluations like cost-utility analysis (CUA) as a decision making tool to improve resource allocation and optimization.

Currently, the number of high-quality CUA studies in this region is quite limited compared to that in Western countries; and, direct CUA result generalization from Western studies could not be substantiated due to differences in socio-cultural, socio-economic, and various other factors. Hence, in order to reduce the need to replicate CUA performed in western countries, it is necessary to identify the factors influencing the generalization of CUA to ensure its role in assessing efficiency of disease management programs in Asian countries.

Within this framework, this thesis is organized to achieve the following objectives:

- To evaluate whether there is any linguistic or cultural barrier in the adaptation of an English health-related quality of life (HRQoL) or utility instrument into a non-English version, as this may influence the results of CUA analyses which require a HRQoL instrument.
- To review the factors found to influence the application, and hence its generalization of CUA in the published literature.
- To explore other new factors that may potentially influence the application of CUA of disease management programs in Asia.

The performed studies revealed the following findings:

- First, minor linguistic and cultural differences did exist between original English version and the translated Chinese version. More importantly, we demonstrated that it was essential to adopt the universalist approach to ensure all important equivalences, namely, conceptual, item, semantic, operational, measurement and functional equivalences were all sequentially and adequately demonstrated during the adaptation process.
- Second, we identified altogether 20 factors from the published literature, which could be used as a reference list by Asian researchers and decision-makers when conducting or adopting CUA analyses. These factors could be further grouped into five categories as treatment-related (duration, efficacy, and frequency), disease-related (severity level, risk level, incidence rate, prevalence rate, disease progression rate and survival length), patient-related (age, gender, race/ethnicity and compliance), cost-related (treatment cost variation, incorporation of indirect cost), and methodology-related factors (discount rate, QALY elicitation method, statistical uncertainty, handling of confounding variables and reliability of data source).
- Third, health psychology-related factors such as response shift and expectancy-value could significantly influence the measurement of HRQoL and utility values, thus impacting CUA results. Comparatively, patient empowerment-related factors such as health literacy and disease knowledge had minimal impact on these values. We also developed and validated two new scales (the generic functional health literacy test and patient's trust in pharmacists scale), which demonstrated high reliability and good construct validity. Their impact on CUA would need further investigation.

In conclusion, we would suggest that if impossible or infeasible to replicate CUA performed in Western countries in an Asian environment, the decision makers need to consider the potential impact of at least the factors we identified when adopting CUA results in formulating policy.

List of Tables

List of Figures

Chapter One

Introduction

1.1 What is disease management?

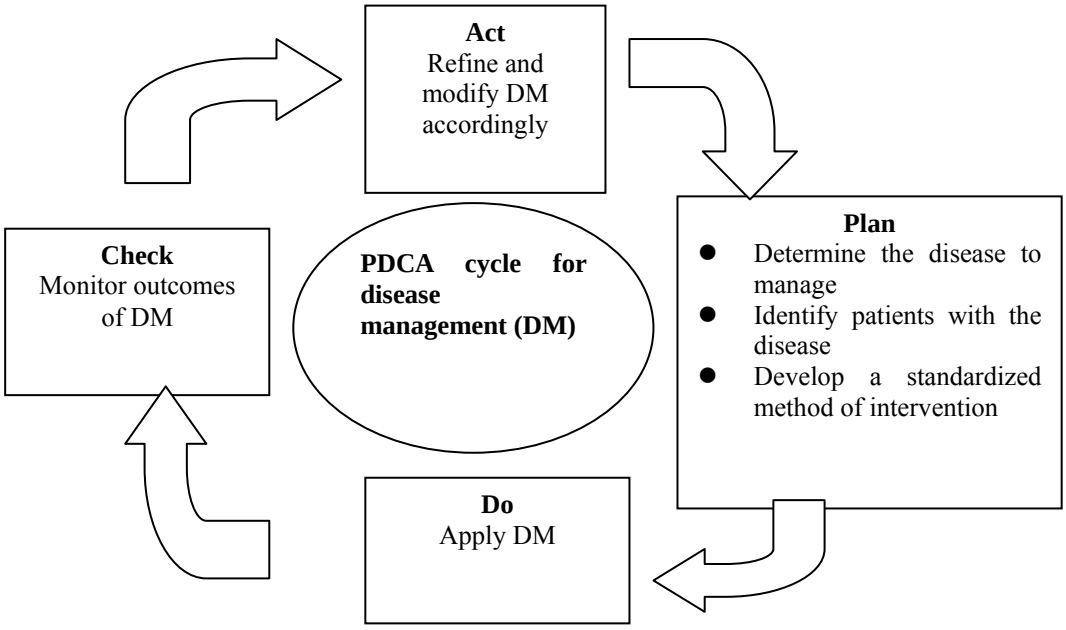
With the ultimate aim to improve the quality of healthcare delivery, disease management is a knowledge-based integrative process intended to continuously maximize the effectiveness of health care delivery at lowest possible expenditures from the perspectives of those who receive, purchase, provide, supply and evaluate it (Couch, 1998; Fritzner et al., Quality and Research Committee, Disease Management Association of America, 2004). In the new millennium, disease management has been increasingly emphasized due to the high prevalence rate of chronic diseases, the pressure of cost containment and the need to improve quality of health care (Fritzner et al., 2004; Fernandes, 2002; Ofman et al., 2004). In Asia, as projected by the United Nations in 2001, the aging population (i.e., the number of people aged 65 and above) in this region will increase by 314 percent, from 207 million in 2000 to 857 million in 2050 (United Nations, 2001). Such a dramatic increase in aging population has been shown to exert ever increasing health and economic burdens on the health care systems across Asia (World Health Organization, 2007). This phenomenon has forced reconsideration about the mode of health care delivery to ensure that acceptable health outcomes can be achieved with affordable expenditure.

Previously, health care was delivered in a so-called "component management system", which often led to increased total treatment cost without the expected improvement in patient outcomes (Todd et al., 1997). Comparatively, in the system of disease management, patients themselves together with other various stakeholders (e.g., policy makers, purchasers, payers, providers, practitioners and product producers, etc.) are included in the value chain of health care delivery with the aims to optimize clinical, economic, humanistic (including quality of life and satisfaction)

outcomes at the lowest possible expenditure (Couch, 1998). This represents a shift from the "piece-meal" approach of health care delivery by individual providers to a more coordinated and streamlined approach aiming to achieve an outcome agreed by all stakeholders. Therefore, the sustainable success of disease management requires refined evidence-based practice guidelines for practitioners, enhanced capability of self-management of patients, valid and reliable outcomes measures by researchers and robust decision-making models for policy makers.

A comparison of the processes of implementing disease management shows that the advocates of disease management approach of health care delivery have borrowed the practices and concepts of total quality management from the business world and applied it to health care with the focus on outcome-based care (Gilmour et al., 1995; Rall et al., 1997; Grol, 2000). As such, the disease management approach also incorporates a feedback loop mechanism to complete the quality improvement cycle known as "Plan-Do-Check-Act" (PDCA) cycle (Figure 1.1). In the Plan-Do-Check-Act loop, assessing disease management could be considered as the procedure of "Check". Therefore, the assessment of disease management not only answers whether disease management is executed properly, but suggests areas for further improvement as well.

Fig.1.1 Plan, Do, Check and Act (PDCA) cycle for Disease Management



1.2 How to assess disease management?

Despite the theoretical benefits of disease management approach over the traditional mode of health care delivery, any disease management program needs to be evaluated to ascertain whether it is realizing its value adding capacity to the health care system. Disease management could be assessed by three major categories of indicators, namely, structure, process and outcomes indicators (Donabedian, 1966). As defined, structure indicators are used to examine physical and organizational properties of health care settings. Process indicators reflect what and how well disease management is executed. Outcomes indicators are applied to show the end results of disease management, such as changes in health status, life expectancy, health-related quality of life (HRQoL) and health care costs (Donabedian, 1980).

Theoretically, outcomes indicators have been regarded as the most comprehensive and representative indicators of the three types. It has been argued that outcomes indicators could directly evaluate the efficiency of disease management

programs, in terms of cost and effectiveness. If designed properly, they are able to reflect all aspects of structure and process, including those that are not measurable or have not been measured yet (Steuten et al., 2006).

However, in reality, it has been found that process indicators were predominantly chosen to assess efficiency of disease management programs, which should have been properly co-examined by outcomes indicators (Mant, 2001). Such phenomenon could be due to the reason that processes are generally easier to measure than outcomes because measuring outcomes requires development and application of new instruments/methods and it usually takes much more time and efforts (Donaldson et al., 2004; Bratzler et al., 2007). Furthermore, outcomes especially from patients' perspectives have not been assessed in a holistic approach, which may lead to potential misinterpretation of the efficiency of a disease management program (Steuten et al., 2006; Ritterband, 2000).

In order to ensure that disease management programs deliver the best outcomes with minimal economic resources, that is, to achieve efficiency in delivery, it is necessary for decision makers to apply systematic economic evaluations to appraise both costs and benefits in a balanced way with no outcome being maximized to the detriment of the other (Gunter, 1999). Based on the guideline issued by the Disease Management Association of America (DMAA) on the "principles of assessing disease management outcomes", it was emphasized that all three types of outcomes, namely, economic, clinical and humanistic outcomes should be evaluated in the assessment of disease management Fritzner et al., Quality and Research Committee, Disease Management Association of America, 2004). One of the most comprehensive

approaches proposed to evaluate outcomes in disease management has been known as ECHO model, which stands for a model evaluating economic, clinical and humanistic outcomes (Reeder, 1995; Gunter, 1999; Kemp, 2006). Clinical outcomes measure the end-points of medical events that occur as a result of disease or treatment and have been used routinely without any controversy in health care settings. Economic outcomes refer to direct, indirect and intangible costs associated with the consequences of medical treatment alternatives or preventions. In the age of cost containment in the realization that resource available for health care is limited, the inclusion of economic outcomes in the assessment of a disease management program is deemed necessary to address the concern of accountability. Humanistic outcomes are consequences of disease or treatment on patient's functional status or quality of life (e.g. physical functioning, social functioning, general health and well-being, and life satisfaction). Additionally, satisfaction with health care services and results of treatment were also an integral part of humanistic outcomes. Comparatively, this is a newer concept of outcome indicator used in measuring the efficiency of a disease management program. Nevertheless, considering that the health care delivery occurring in any disease management program can be conceptualized as a humanistic exchange between the providers and the receivers (patients in this case), it would be logical to include the impact from the perspective of the receiver. Hence, the ECHO model of assessing outcomes in disease management has been gaining increasing acceptance since its proposal.

Since the adoption of ECHO model, apart from continuous efforts in seeking better clinical indicators, there have been increasing research endeavors in the development and validation of patient-reported outcome measures (e.g. HRQoL,

functioning and satisfaction measures) to evaluate humanistic outcomes, which could further contribute to the accurate assessment of either benefit or effectiveness in economic evaluations (Kind, 2001; Korolija, 2007).

The other equally important issue in the application of ECHO model to assess disease management is to enhance the robustness of economic modeling, which has been heavily pursued by developing and refining guidelines, consolidating individual approaches, and seeking appropriate application of these approaches for various purposes (Fleurence et al., 2007; Mason et al., 2006; Inadomi, 2004; Siegel et al., 1997).

Nevertheless, for a brief summing up, the outcome indicators from the ECHO model are being increasingly used in economic evaluation of disease management program worldwide and signify a philosophical shift in assess efficiency of health care delivery.

1.3 What is the role of cost-utility analysis (CUA) in assessing disease management?

Currently, there are altogether four basic economic evaluations used to assess disease management programs, namely, cost-minimization analysis (CMA), cost-benefit analysis (CBA), cost-effective analysis (CEA) and cost-utility analysis (CUA) (Johannesson,1996). Collectively speaking, all of these four economic evaluations generate results by comparing both the cost and health outcomes of two or more interventions using one formula, with difference in costs presented in the nominator and difference in health outcomes in the denominator.

As shown in Table 1.1, the four economic analyses share one commonality, that is, to measure costs in monetary value. Except for CMA which assumes the health outcomes to be the same and thus requires no measurement, all the other three types of economic evaluations measure health outcomes in different ways. CBA measures health outcomes in monetary value, which has generated argument that it is inappropriate and difficult to place a dollar value on human life. Comparatively, both CEA and CUA measure benefits in non-monetary units such as quantity or quality of life, which are more acceptable and make clinical sense in the assessment of health care (Muenning, 2002).

Table 1.1 The four basic economic analyses to assess disease management programs

Type of analysis	Costs	Outcomes	Comments
Cost Benefit Analysis (CBA)	In monetary units	In monetary terms	Less commonly used as it is difficult to quantify human life in monetary units
Cost Minimization Analysis (CMA)	In monetary units	Assuming equivalence in outcomes	To identify the cheapest program when the outcomes are assumed to be the same.
Cost-effectiveness analysis (CEA)	In monetary units	In physical or natural units	Mainly used to compare programs across the same disease/treatment
Cost-utility analysis (CUA)	In monetary units	In utility-weighted life years	Allows comparison across different diseases/treatment

However, more specifically, the unit of health outcome measures also differs between CUA and CEA, whereby CUA measures health outcomes in utility-weighted life years and computes a cost per utility-measure ratio for comparisons across different interventions (Gerard, 1992). Consequently, CUA has been recommended as the preferred form of reporting health economic evaluation results by the Panel on

Cost-Effectiveness in Health and Medicine (Siegel et al., 1996).

In CUA, the utility-weighted life years could be either calculated as Quality-Adjusted Life Years (QALYs) or Disability-Adjusted Life Years (DALYs) (Gold, 2002; Torrance, 1997; Sassi, 2006).

Congruent with the purpose of maximizing allocative efficiency in disease management, the use of such a standardized index in CUA would allow not only comparison of cost-effectiveness among interventions for the same disease or condition, but also comparison across interventions for different diseases or cost-effectiveness of different health service programs. This is a distinct advantage over the use of CEA where the major concern would be that of technical efficiency.

Regarding the standard index commonly used in CUA, besides quantity of life (life expectancy), QALYs also incorporate quality of life (health utilities), which has been shown to be a very useful and important humanistic endpoint to assess the effectiveness of treatment, particularly in chronic diseases where mortality is not the major issue or when the primary purpose of the intervention is palliative rather than curative (Burger, 2003; Merhrez et al., 1989; Raisch, 2000). Health utilities could be either elicited by direct measurement using Time Trade-Off (TTO), Standard Gamble (SG) and Rating Scale (RS) (Morimoto et al., 2002), or by indirect measurement using utility-based Health-related Quality of Life (HRQoL) instruments such as SF-6D (Brazier et al., 2002), EQ-5D (The EuroQol Group, 1990), Health Utilities Index Mark 2 (HUI2) (Torrance et al., 2002) and Mark3 (HUI3) (Feeny et al., 2002).

In comparison, DALYs can be viewed as a form of unequally weighted QALYs, which assign different weights to different age groups when quantifying the burden of disease (Murray, 1994). Yet, all other non-health characteristics of an individual are arbitrarily ignored, which is based on Murray and Acharya's ethical principle that all other factors should play no part in deciding health care priorities (Murray, 1997). Due to the increased complexity in DALYs calculations and widely challenged conceptual and technical soundness, DALYs are much less adopted than QALYs as a standardized form to measure benefits in CUA (Fox-Rushby, 2001), but used more commonly in assessment of burden of illness.

Hence, this near univocal acceptance of QALY as the outcome indicator in CUA should promote the application of CUA in evaluating disease management at least theoretically.

1.4 Why explore factors influencing the application of CUA to assess disease management among Asian patients?

Despite heterogeneities, health care systems across Asia are facing unprecedented and unparalleled increase in health care expenditure, due to the aggregated impact of rapid growth in aging population, rising prevalence in chronic diseases and increasing demand of better quality of health care (Clark, 2004; Cheah, 2001; East West Center, 2002). For most of the countries in the region (as shown in Table 1.2), although the total expenditure on health as percentage of Gross domestic product (GDP) is seen to be more or less stable, yet when GDP growth is factored in, the per capita total expenditure on health is found to be subject to continual growth at relatively high rates (World Health Organization, 2007). With the rapidly ageing of populations in the

region as previously mentioned, this trend is likely to continue if not escalated in the next few decades. Some sort of systematic approach, such as the adoption of disease management would be necessary to ensure that the increased health care expenditure is delivering the required outcomes. Hence, economic evaluations like CUA would play a significant role in improving the rationality of disease management in terms of resource allocation and optimization.

Probably due to the fact that disease management assessment is still a relatively emerging area in Asia, the number of high-quality CUA studies is quite limited in this region compared with that in Western countries. Direct generalization of CUA results from studies in the West could not be substantiated, as variations in the CUA results have been found even across different locations in those countries (Sculpher et al., 2004). Hence, it is deemed necessary to perform CUA based on local populations and health care settings to generate more accurate appraisal of the locally implemented disease management programs.

However, for the successful application of CUA to assess the efficiency of disease management programs in Asian countries, there are a number of factors either as barrier(s) or concerns that need to be solved or addressed.

Table 1.2 National Health Accounts of Selected Asian Countries*

Country	Total expenditure on health as % of Gross domestic product (GDP)					Per capita total expenditure on health at average exchange rate (US$)					
	2000	2001	2002	2003	2004	2000	2001	2002	2003	2004	Average annual Growth rate (2000-2004)
Cambodia	5.9	6.6	7.1	7.3	6.7	17	19	22	24	24	8.5%
China†	4.6	4.6	4.8	4.8	4.7	44	48	54	62	71	12.9%
Democratic People's Republic of Korea‡	3.6	3.5	3.5	3.5	3.5	17	18	21.2	21.2	21.2	6.1%
India	4.3	4.5	4.8	4.9	5	19	21	23	27	31	12.9%
Indonesia	2.3	2.7	2.8	2.9	2.8	18	20	25	31	33	16.4%
Japan	7.6	7.8	7.9	8	7.8	2 828	2 558	2 450	2 694	2 831	0.3%
Lao People's Democratic Republic	3.2	3.3	3.3	4.5	3.9	11	11	11	15	17	13.5%
Malaysia	3.3	3.7	3.7	4.2	3.8	130	138	146	177	180	8.8%
Mongolia	7.9	8.2	8	6.5	6	30	33	35	32	37	5.9%

Table 1.2 (Continued)

Country	Total expenditure on health as % of Gross domestic product					Per capita total expenditure on health at average exchange rate (US$)					
	2000	2001	2002	2003	2004	2000	2001	2002	2003	2004	Average annual Growth rate (2000-2004)
Myanmar	2.1	2.1	2.3	2.2	2.2	3	3	3	4	5	10.9%
Philippines	3.5	3.2	3	3.4	3.4	34	30	29	34	36	2.0%
Republic of Korea	4.8	5.4	5.3	5.5	5.6	518	553	607	703	787	11.1%
Singapore	3.6	4.3	4.2	4.2	3.7	820	888	894	921	943	3.6%
Thailand	3.4	3.3	3.7	3.5	3.5	68	62	75	80	88	7.3%
Viet Nam	5.3	5.5	5.1	5.2	5.5	21	23	22	25	30	9.8%

*: Source: World Health Organization. NHA ratios and per capita levels (Excel). Available from: http://www.who.int/nha/country/en/index.html. Accessed on January 27, 2008.

†: The estimates do not include expenditures of Hong Kong and Macao Special Administrative Regions.

‡: Exchange rate was changed from 2.15 Won in 2001 to 152 Won in 2002. For comparison, the figures in the table were all based on the exchange rate of 2.15 Won.

First and foremost among these factors is the linguistic barrier. Although a variety of reliable disease-specific and generic HRQoL and utility measures were available for use in CUA, most of them have been developed in Western countries (in particularly English speaking countries) in languages and contexts that are different from those of Asian countries. Ideally, it would be necessary to develop HRQoL instruments for use in different Asian countries and cultures, but the resource requirements would be overwhelming and the comparability of results across countries using country-specific instruments would be problematic. As a trade-off, in order to avoid the time and efforts of developing a new instrument, translating and adapting suitable ones in an Asian population is a wiser choice but it requires comprehensive validation process to ensure their reliability and validity.

Secondly, in order to perform high-quality CUA to assist decision-making in uncertainties, it is important to generate a list of potential factors that need to be incorporated into sensitivity analyses to examine their impact on the robustness of CUA results.

Thirdly, in view of additional impact of differences in socio-cultural, socio-economic and socio-epidemiological differences between the East and the West, it is necessary and important to explore new factors that could potentially influence CUA to enhance its comprehensiveness and robustness in assessing disease management among Asian patients.

These are some of the pertinent concerns that need to be addressed properly in order to facilitate the proper application of CUA in disease management assessment in

Asian countries. The studies performed and reported in this thesis are to address these concerns.

In the studies presented in this thesis, Singapore was selected to provide a demonstrative study population within the Asian region due to three major reasons which are listed as follows:

(1) Singapore is one of the leading countries in Asia adopting disease management (Cheah, 2001), which has greater needs to have robust economic analysis like CUA studies to be in place;

(2) Singapore has a multiethnic and multilingual population of Chinese (76% of the total population), Malays (14% of the total population) and Indians (8% of the total population), most of whom may share similar socio-cultural background and values with their counterparts in China, Malaysia and India; hence, with this unique position, the results generated from Singapore are expected to serve as better references for researchers in the Asian region to further investigate the factors influencing the application of CUA in their own countries; and

(3) Singapore is among the most westernized of all the Asian countries, but at the same time maintains a very strong root in traditional Asian values and culture. Hence Singapore would provide an ideal transitional site for studying of the various factors impacting on CUA. Analogous to a filtering system to save energy and efforts, any factors found not to be impacting on CUA in Singapore would be unlikely to be important in many other Asian countries, while factors found to be important could be tested further.

1.5 Research objectives

In general, the current thesis aims to provide a reference list of the factors that may have potential impact on the application of CUA to assess disease management in Asian countries.

Specifically, in line with the barrier and concerns as mentioned in Section 1.4, the subsequent eight chapters were designed to answer the following research questions:

1. Is there any linguistic or cultural barrier in the adaptation of an English HRQoL or utility instrument into a non-English version among Singaporeans? As generic HRQoL and utility measures such as EQ-5D and SF-6D have already been adapted into Chinese, Malay and Tamil for EQ-5D (Luo et al., 2003; Wee et al., 2007) and into Chinese for SF-6D (Wee et al., 2004) in Singapore, hence the efforts were focused on the demonstration of adaptation of disease-specific HRQoL instruments, which had been found to be more responsive and accurate to measure HRQoL or utility changes compared with those generic measures (Mishoe et al., 2001; Hart et al., 2007; Eurich et al., 2006) Furthermore, items in the disease-specific HRQoL instruments were recommended to be used as a reservoir for selection of different health profiles for the direct elicitation of utility values by SG, TTO or RS (Buxton et al., 2007; Krahn et al., 2007; Casey et al., 2006), the validity and feasibility of which have been shown among a multiethnic Asian population in Singapore (Wee et al., 2008). The studies as presented in Chapter 2 would illustrate as an example of how to translate and validate the English Audit of Diabetes-Dependent Quality of life (ADDQoL) into a Chinese version based on a systematic approach that could evaluate all six types of equivalences, namely, conceptual, item, semantic, operational, measurement and functional equivalences.

2. What are the factors that have been found to influence the application of CUA in the published literatures so far? A qualitative literature review on such factors would be presented in Chapter 3. Factors identified from the literature review could be used by researchers or decision makers in Asia as a fundamental reference list to verify and evaluate their potential impacts on future Asian CUA studies. Moreover, the availability of such list also serves as a stepping stone for us to explore other new factors that would potentially influence the CUA in the Asian populations and contexts.

3. Are there any other new factors that may potentially influence the application of CUA of disease management programs in Asia? In this thesis, patient empowerment-related and health psychology-related factors would be the two major categories to be explored.

As patient empowerment strategies were found to be another key factor that might enhance the effectiveness and efficiency of disease management programs (Thiel de Bocanegra and Gany, 2004; Day, 2000; Anderson, 1996), therefore it was deemed necessary to explore the correlation between patient empowerment factors such as disease knowledge and health literacy with health utility values. If a significant relationship between those factors with health utility values is identified, the magnitudes of impact of these two factors on CUA results would also be discussed at the end of each chapter.

Correspondingly, the study presented in Chapter 4 aims to investigate the impact of disease knowledge using diabetes knowledge as an example, while the study in Chapter 5 focuses on the correlation between health literacy with health utility values,

using patients with rheumatic diseases as an example.

In the study presented in Chapter 5, a widely used word recognition test, called the Rapid Estimate of Adult Literacy in Medicine (REALM), was used to assess patients' health literacy levels. However, there have been debates around whether REALM could assess patients' comprehension capabilities of the medical information, or the functional health literacy levels of the patients (Friedman and Hoffman-Goetz, 2006). Consequently, it actually generated research interests to develop and validate a few functional health literacy tests by researchers in the West [e.g. the Newest Vital Sign (NVS), the Test of Functional Health Literacy in Dentistry (TOFHLiD) and the Test of Functional Health Literacy among Adults (TOFHLA)] (Parker et al., 1995; Weiss et al., 2005; Gong et al., 2007).

However, due to the lack of content validity of those available functional health literacy measure for use among patients or general public in Singapore, a generic functional health literacy test was therefore developed and validated (presented in Chapter 6) for further investigation of the impact of functional health literacy on CUA in future studies.

Chapter 7 to Chapter 9 are dedicated to the studies of three health psychology-related factors, namely, response shift, health preference and trust in pharmacists. The study in Chapter 7 uses the total knee replacement patients as an example to explore the impact of response shift on the longitudinal measurement of health utilities, as the presence of response shift may directly affect the robustness of CUA results.

The study in Chapter 8 presents a comprehensive psychological model named as "expectancy-value model" to better explain the differences in health preferences to facilitate the understanding of health utility values across different populations. Last but not least, the study in Chapter 9 focused on the development and validation of a new scale to measure patient's trust in pharmacists, as trust in pharmacists may influence patient's satisfaction and adherence to disease management programs. The availability of such scale could therefore contribute to the further investigation of its impact on future CUA studies in Asia.

Chapter Two

Translating and culturally adapting the English version of Audit of Diabetes-Dependent Quality of Life (ADDQoL) into Chinese-- An exploratory study based on the universalist approach

2.1 Introduction

Diabetes mellitus has become one of the most daunting public health problems in the world, because of its alarmingly increasing prevalence, significant impairment on patients' quality of life (QoL), and tremendous burden on healthcare resources. The global number of individuals with diabetes in 2006 was estimated to be 180 million, a figure projected to be more than doubled by 2030 [World Health Organization (WHO), 2006]. According to the estimate by the International Diabetes Federation (IDF) in 2006, China was ranked as the country with the highest prevalence in diabetes with a disease population of 39 million (IDF, 2006). With such high prevalence rates internationally, diabetes and its related complications impose significant economic consequences on individuals, families, health systems and countries. WHO estimates that over the period of 10 years from 2006 to 2015, in China alone, a loss up to $ 558 billion in foregone national income would be incurred due to heart disease, stroke and diabetes (WHO, 2006).

Singapore, another Southeast Asian country, also has a higher diabetes prevalence of 8% among its predominant Chinese ethnic group compared with other parts of the world (Ministry of Health of Singapore, 1998). With such a high prevalence, diabetes is causing significant morbidity and mortality in Singapore, and the implementation of disease management programs for diabetes would be a cost-effective way to reduce health and economic burdens and improve patients' QoL, in particular their health-related quality of life (HRQoL).

HRQoL is a subset of QoL, which describes patient–perceived functional effect of an illness and its consequent therapy. With the international trend of ageing of the

populations and the resultant increasingly prevalence of chronic diseases requiring often life-long treatment, the adoption of HRQoL as an outcome indicator in addition to the conventional biochemical outcome indicators has been gaining acceptance clinically. Similarly, due to the chronic nature of diabetes, HRQoL has been increasingly used as a supplementary outcome measure in addition to the traditional biomarkers like HbA1c values. Furthermore, HRQoL has been incorporated into utility assessment and has become an important parameter for assessing the cost-effectiveness of such programs either by CEA or CUA. With regards to HRQoL measurement, the major challenge to researchers, clinicians and decision-makers alike is to find a suitable HRQoL instrument that is both reliable and sensitive.

For diabetes, there are several HRQoL instruments available (Luscombe, 2000; El Achhab et al., 2008; Brazier et al., 1998; EuroQol Group, 1990). Among these, the Audit of Diabetes-Dependent Quality of life (ADDQoL) is a valid and reliable diabetes-specific HRQoL measure originally developed in U.K. It has two distinguished and unique features: one is to allow patients to indicate which aspects of life apply to them by using the "not applicable" (N/A) options; the other is the application of importance ratings of each domain so as to give a weighted score in the end (Bradley et al., 1999). Compared with other generic measures like EQ-5D and SF-6D, these two features make the ADDQoL both a more sensitive HRQoL instrument and a valuable candidate to be further developed for utility assessment in CUA of diabetes programs.

The English version of the ADDQoL has been culturally adapted for use in Singapore without any modification and the equivalence between the adapted and the

original version was also demonstrated recently (Wee et al., 2006). However, to date, a Chinese version of ADDQoL is not available. In Singapore itself, about 32% of the local ethnic Chinese is monolingual in Chinese (Singapore Census of Population, 2000). Given the large number of monolingual Chinese-speaking diabetic patients in the world and the importance of HRQoL and utility measurement to evaluate cost-effectiveness of diabetes management programs, a culturally adapted and validated Chinese version of ADDQoL would contribute significantly to the management of DM in Chinese patients.

A universalist approach to the cross-cultural adaptation of HRQoL instruments proposed by Herdman et al (1998) suggests that conceptual equivalence and item equivalence should be examined and demonstrated before we start to translate a questionnaire into the target language. The universalist model of cross cultural adaptation criticized a commonly used approach, where translation is completed first and then *post hoc* analysis is performed to demonstrate equivalence especially the measurement equivalence. Such commonly used approach overlooks the evaluation of the conceptual and item equivalences, which is of fundamental significance in identifying any potential cultural barriers in the cross cultural adaptation process.

According to the universalist approach, six types of equivalence should be investigated in sequence as follows:

(1) Conceptual equivalence to investigate which domains are important to the concept in the target culture and the relationships between them, which can be achieved by reviewing local literature, consulting experts in the target culture and

discussing with target group;

(2) Item equivalence to examine critically the items used to tap those domains as the relevance of items may vary across cultures which can also be achieved by literature review, expert judgment and assessment by target population;

(3) Semantic equivalence to ensure that any translation which takes place leads to semantically equivalent items with the recommended translation process is to be done according to the following steps: initial discussion with the developer about the underlying concept (this step should be completed in the phase of "conceptual equivalence"), forward translation, cognitive debriefing, backward translation, cognitive debriefing, harmonization review, feedback by developer, revision, proofread and approval of final version by the developer (Acquadro et al., 2004);

(4) Operational equivalence to ensure that the measurement methods used are appropriate to the culture in question which can be investigated by using similar methods as mentioned in "item equivalence";

(5) Measurement equivalence to examine the outcome of the process in terms of instrument behavior; reliability, responsiveness, construct validity (convergent and divergent validity, known group validity) tests are often used; and

(6) Functional equivalence to summarize the above-mentioned types of equivalence.

The purpose of the study was to translate and culturally adapt the English-ADDQoL into Chinese for use in Singapore with the universalist approach. As the English-ADDQoL was previously adapted in Singapore without any modification, therefore the one used for adaptation in our study was actually the same as the original U.K. version. The Chinese version developed in this study was aimed to be used in Singapore first. The possibility of adaptation to other Chinese-speaking

population could be assessed in future studies.

As for the tests of construct validity, the following 4 *a-priori* hypotheses were generated based on literature review:

(A) Convergent and Divergent validity

1. An assumption that the "Present HRQoL" score will correlate moderately with the EQ-5D utility, SF-6D and Visual Analogue Scale (VAS) scores. The assumption is based on literature reports that disease specific instruments correlated moderately with utility-based instruments (Luo et al., 2003; Revivki and Kaplan, 1993).
2. The ADDQoL mean weighted score will correlate moderately with "HRQoL without diabetes" and correlate weakly with "Present HRQoL" score (Nunnally, 1978).

(B) Known group validity

1. Subjects who are more depressed (as shown by the score in the mental health in SF-6D) will have poorer ADDQoL scores (Paschalides et al., 2004)
2. Subjects who have better family functioning [as shown by higher score in Family Functioning Measure (FFM)] will have better ADDQoL mean weighted scores (Sawyer et al., 2001).

2.2 Methods

2.2.1 Subjects and study design

This study was carried out in two phases. The first phase was to use the universalist approach in translating and culturally adapting the English version of

ADDQoL into a Chinese version. The second phase of the study involved pilot testing the adapted Chinese version and evaluation of its equivalence with the English version.

During the first part of the study, as suggested by the universalist approach, steps were taken to investigate conceptual equivalence, item equivalence, semantic equivalence and operational equivalence, through which translation was integrated. Conceptual, item and operational equivalence was assessed by local literature review, expert judgment and cognitive debriefing among target subjects. Semantic equivalence was studied according to the recommended translation procedure which will be described in the translation part of the methodology. Two local bilingual (Chinese and English) clinical experts in diabetes were involved in the judgment.

Five native Chinese-speaking diabetic patients were recruited for the cognitive debriefing during the whole process of the first part. They were members of Diabetes Society of Singapore (DSS) who attended two government polyclinics. These recruited patients differed in their gender, type of diabetes (diagnosed by physician) and treatment method to better represent the diabetic patients in Singapore. Of the 5 patients, there were 1 male and 1 female patient each with Type-1 diabetes; the other patients consisted of 2 male and 1 female Type-2 diabetics who were undergoing medication, diet control, and insulin injection respectively. Their consents to participate ion the study were sought through a consent form before the interview.

After the above four kinds of equivalence were demonstrated, measurement equivalence was investigated by a pilot cross-sectional study during the second phase

of the study. It was undertaken at the World Diabetes Day 2004 commemorative event organized by DSS at the exhibition hall of a major shopping center in Singapore on Oct 17, 2004. Participants were approached by research assistants to identify whether they were eligible for the study. Inclusion criteria were English-speaking or Chinese-speaking Singaporean diabetic patients (type 1 or type 2) aged 21 and above, who were able to complete questionnaires without any assistance. All subjects who agreed to participate in the study were asked to sign a written informed consent. Then they were asked to complete a booklet containing a demographic datasheet, diabetes knowledge test sheet (the results of which were used and reported for another study), FFM and a battery of HRQoL instruments including ADDQoL and two other generic HRQoL measures (SF-6D and EQ-5D).

A week later, those who had indicated in the consent form that they were willing to take part in the retest, received a similar set of questionnaires, excluding the demographic sheet by mail together with a stamped returned envelope. A reminder call was given to patients if they failed to return the mail 2 weeks after the administration.

2.2.2 Translation

2.2.2.1 Initial discussion with the developer

The developer was contacted and asked to provide descriptions of ideas behind the language used in original ADDQoL.

2.2.2.2 Forward translation (FT)

In the first stage, a preliminary Chinese version from the original English

ADDQoL was developed. Two Singaporean translators, both fluent in writing and speaking Chinese and English, did the initial forward translation independently. After the two forward translations were completed, the translators met with the project coordinator to produce a preliminary consensus version of the ADDQoL in Chinese (C-ADDQoL FT). A forward translation report was produced at the end of the meeting.

2.2.2.3 Cognitive debriefing

After the forward translation was produced, cognitive debriefing was carried out. Conceptual equivalence, item equivalence and operational equivalence were discussed with participants during the interview. Participants were asked about their view of health and QoL (conceptual equivalence), relevance of each item to themselves (item equivalence). Besides, they were asked whether items were difficult to understand or to answer (for translation purpose). Operational equivalence was assessed in two steps: first, participants were asked whether they were able to finish the self-administered ADDQoL; then, they were required to complete the questionnaire and discuss with interviewer about the appropriateness of the questionnaire format, instructions and mode of administration. The cognitive debriefing report was generated after the debriefing. It included items that had been changed from the tested version and the reason for such a change. The edited version of the ADDQoL FT was produced and subjected to backward translation (BT).

After backward translation has been completed, another cognitive debriefing was carried out before a preliminary version of Chinese ADDQoL (C-ADDQoL pre) was produced.

2.2.2.4 Backward translation (BT)

After the first cognitive debriefing, another two bilingual Singaporean translators were recruited to back translate the C-ADDQoL pre from the forward translation into English independently. Once the back translation was completed, the project coordinator met up with the translators to discuss the problems during the translation. The Chinese-ADDQoL Backward Translated version (C-ADDQoL BT) was produced after all discrepancies were dealt with. Next, a backward translation report was compiled to indicate the differences between the two back translations and the differences between the latter BT version and the original one was noted and recorded. This C-ADDQoL BT was then subjected to the second cognitive debriefing.

2.2.2.5 Harmonization review, feedback and approval by original developer

Finally, all the translators met up with coordinator and discussed to compose the C-ADDQoL pre by settling down any discrepancy or disagreement before submitting the version and all the forward translation, backward translation, cognitive debriefing reports to the original developer for a final review. The original developer then did a review to ensure the authenticity of the translation. After the approval had been obtained, the final version of the C-ADDQoL was produced.

2.2.3 Instrument

2.2.3.1 ADDQoL

ADDQoL-13.3.03 is a 19-domian disease-specific instrument designed to measure individual's perception of the impact of diabetes on their quality of life (See Appendix 2.1). It begins with two items assessing "present HRQoL" and "HRQoL without diabetes". The former is a measure of general HRQoL, while the latter is a

measure of diabetes-specific HRQoL. The subsequent items are related to the 19 domains of physical functioning, symptoms, psychological well-being, social well-being, role activities, personal constructs, etc (Bradley et al., 1999). Impact of each domain is scored on a 5-point scale (from -3 to 1) and the corresponding importance is rated on a 4-point scale (0 to 3). The weighted score of each domain is calculated by multiplying the impact and importance rating (-9 to +3).

Out of the 19 domains, there are five with N/A (not applicable) options. These N/A items were scored as missing values. As a result, domains that are not important to respondent are excluded from the mean ADDQoL weighted score. Lastly, the mean weighted score is converted to a final score on a 0-100 scale. A higher score indicates a better HRQoL.

2.2.3.2 EQ-5D

The EQ-5D questionnaire is a 5-item (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) questionnaire for respondents to self-classify and rate their health on the day of administration of the instrument (Brooks, 1996; Rabin and de Charro, 2001). For each item, there are three response levels (with no problem, with some problems, with extreme problems). Theoretically, 243 unique health states could be identified by the descriptive system of the instruments.

Scoring methods have been developed to assign each of these health states a utility score, in which 1 represents full health (no problem with all 5 items) and 0 represents being dead (Dolan, 1997; Dolan and Roberts, 2002). The range of the final score is from -0.594 to 1.00. The validity and equivalence of the EQ-5D in English

and Chinese versions were previously demonstrated in a sample of rheumatic patients in Singapore (Luo et al., 2003).

2.2.3.3 SF-6D

The SF-6D is a 6 dimensional health classification system assessing physical functioning, role limitations, social functioning, pain, mental health and vitality, with 4 to 6 level per dimension (Brazier et al., 1998; Brazier et al., 2002). An SF-6D health state is defined by selecting 1 level from each dimension. The SF-6D score is scaled from 0.26 to 1.00 continuously, with 0.26 representing the worst health state (all dimensions being at the worst level) and 1.00 representing full health (all dimensions being at full functional level).

The validity and equivalence of the SF-6D in English and Chinese versions were previously demonstrated in a population-based study in Singapore (Wee et al., 2004).

2.2.3.4 Family functioning measure (FFM)

The family functioning measure is a 3-item instrument designed to measure the quality of interaction among family members (Sherbourne and Kamberg, 1992). It has a 5-point Likert scale, with higher scores reflecting better family functioning.

Several studies have shown that family functioning was associated with HRQoL scores (Thumboo et al., 1999; Sawyer et al., 2001). The validity for use among Singaporeans had been demonstrated previously (Thumboo et al., 1999).

2.2.4 Statistical analysis

Data collected from the second phase of the study were entered into an Excel spreadsheet (Microsoft Corporation, Redmond, Washington) and analyzed using the Statistical Package for the Social Sciences, version 12.0 (SPSS Inc, Chicago, Ill). Subjects with ≥6 missing items in ADDQoL or ≥ 1 missing value in EQ-5D or SF-6D were excluded from all analyses.

Group comparisons were made using non-parametric tests with Mann-Whitney U tests for quantitative data and Chi-square tests for qualitative data. For reliability tests, Cronbach's α and the intraclass correlation coefficient were calculated for internal consistency and test-retest reliability, respectively. Response distribution for each item, use of N/A options and item ranking (weighted, unweighted and change in order of ranking) were compared between two language groups as responsiveness tests, during which descriptive analyses and group comparisons were carried out. Spearman's correlation coefficients were used to investigate construct validity (convergent and divergent validity, known group validity) by correlation analyses among ADDQoL ("present HRQoL" score, ADDQoL mean weighted score, "HRQoL without diabetes" score), SF-6D and EQ-5D scores. Statistical significance was set at $p<0.05$ for the data analysis and $p<0.01$ or <0.001 was also reported where applicable.

2.3 Results

2.3.1 Characteristics of the subjects

Altogether 56 eligible English-speaking patients and 32 eligible Chinese-speaking patients participated in the study. During data analysis, 14 English-speaking subjects and 6 Chinese-speaking subjects were dropped due to incomplete responses as defined previously, leaving 68 subjects (42 English-speaking subjects and 26 Chinese-speaking subjects) with complete responses for analysis. Table 2.1 shows respondents' characteristics and scores of FFM and HRQoL measures.

Respondents who completed the Chinese-ADDQoL were more likely to be unemployed ($p<0.01$), have received fewer years of education ($p<0.05$), lived in public housing ($p<0.01$) and poorer family functioning ($p<0.01$). Although statistical significance was not demonstrated, English-speaking patients appeared to have higher scores in EQ-5D, SF-6D and ADDQoL .

Table 2.1. Characteristics of study subjects

	N (%) unless stated		
	English (n=42)	Chinese(n=26)	p-value
Age [Mean (SD)]	53.5 (9.64)	55.8 (11.09)	0.479
Female	18 (45.0)	12(46.2)	0.068
Chinese ethnicity	36(85.7)	26 (100)	
Years of education			<0.05
≤6	4 (10)	22 (84.6)	
7-10	18 (45)	2 (7.7)	
≥11	18 (45)	2 (7.7)	
Employed	23 (54.8)	4 (15.4)	<0.01
Smoking	3 (7.1)	3 (11.5)	0.535
Presence of acute medial conditions*	23 (54.8)	15 (57.7)	0.813
Presence of chronic medical conditions other than diabetes†	27 (64.3)	14 (53.8)	0.393
Presence of diabetes complications‡	17 (40.5)	13 (50.0)	0.442
Housing type			<0.01
Public			
Lower cost	1 (2.4)	16 (61.5)	
Regular	32 (76.2)	6 (23.1)	
Private	9 (21.4)	4 (15.4)	
Type of diabetes			0.287
Type I	13 (31.0)	5 (19.2)	
Type II	29 (69.0)	21(80.8)	
Mean Family function score (SD)	63.5 (16.21)	47.0 (17.46)	<0.01
Median (Interquartile range) EQ-5D utility	1.00 (0.81, 1.00)	1.00 (0.78, 1.00)	0.331
Median (Interquartile range) SF-6D	0.80 (0.70, 0.95)	0.77 (0.65, 0.94)	0.303
Median (Interquartile range) weighted ADDQoL	-3.05 (-4.38, -1.26)	-4.31 (-6.10, -2.24)	0.238
Median (Interquartile range) converted ADDQoL	49.57 (38.53, 64.47)	39.15 (25.00, 56.40)	0.241

*: Acute medical conditions included running a nose, sore throat or cough, vomiting or diarrhea, headache lasting more than one day, sleeping problems and body injuries.
†: Chronic medical conditions other than diabetes included hypertension, heart disease, stroke, asthma or other lung disease, cancer, rheumatism, back pain or other bone or muscle illness, mental illness, other illness like kidney problems on dialysis.
‡: Diabetes complications included eye disease, foot problems, kidney disease, heart disease, stroke, erectile dysfunction and neuropathy.

2.3.2 Conceptual equivalence

The perception of HRQoL among Chinese-speaking Singaporeans has been well demonstrated by several local studies (Thumboo et al., 2002; Luo et al., 2003; Wee et al., 2004). Relevance of 19 domains to the Chinese-speaking Singaporean diabetes patients was confirmed by two local diabetes experts and respondents during the cognitive debriefing. As domains are weighted in ADDQoL, the importance of the domains is allowed to vary between two cultures.

2.3.3 Item equivalence

Initial qualitative examination of relevance of items was carried out using the same procedures as described in the "conceptual equivalence". The results suggested that items can be used in the Chinese version without modification other than translation (See Appendix 2.2 for the Chinese-ADDQoL). Quantitative investigation of item equivalence will be reported in the "measurement equivalence" section later.

2.3.4 Semantic equivalence

No significant problem surfaced during the translation process except for a few minor lexical issues. Among them, one issue was that sometimes an English word or phrase possesses multiple explanations in Chinese. During forward translation, the word "close" (in the context of "close personal relationship") and another phrase "working life" were translated into two different Chinese terms respectively. Similar problems were also encountered in the backward translation for the word "physically" and the phrase "living condition". Such problems were then resolved by selecting the most appropriate Chinese term, which could convey the original meaning indicated by the developer.

Some other discrepancies were due to the difference in word order between the Chinese and English expressions. In the first cognitive debriefing, all five respondents found the translation of "much easier" for the impact scale option to be awkward. Nevertheless, it was used in order to maintain the thematic meaning (consistent word order with other impact scale options). In the second cognitive debriefing, the same problem was brought up again. After much lengthy discussion, in the end the translation was deemed suitable as consistency with the English version should be adhered to.

No major problem arose during the harmonization review. The C-ADDQoL pre together with all the necessary reports were then sent to the original developer for final review. Since the version was approved without any change, it was thus adopted as the final C-ADDQoL.

2.3.5 Operational equivalence

The previous study of the English-ADDQoL adaptation in Singapore and other studies of similar questionnaires used among Singaporeans have shown that the format of ADDQoL and the mode of self-completion would be feasible in Singapore (Wee et al., 2006; Thumboo et al., 2002; Luo et al., 2003; Thumboo et al., 1999; Koh et al., 1998). Apart from literature review, this was also confirmed by researchers in outcomes research field in Singapore. Additionally, respondents were able to complete the questionnaire without any trouble during cognitive debriefings, which predicted the success of the testing methods.

2.3.6 Measurement equivalence

2.3.6.1 Reliability of the Chinese-ADDQoL

Cronbach's alpha indicates how much the items on a scale are measuring the same underlying dimension, that is, a measure of internal consistency (Pilot, 1996). In this study, the Cronbach's alpha for the overall items is 0.94. When two items of "self-confidence" and "financial situation" were deleted individually, the Cronbach's alpha reduced to 0.93. However, if the item of "freedom to drink" was deleted, the alpha increased to 0.95. As for the separate deletions of other items, the Cronbach's alpha remained at 0.94 (See Table 2.2). All of the above coefficient alphas suggested excellent reliability, because according to the rule of thumb that applies to most situations, a value exceeding 0.9 would demonstrate excellent reliability (George and Mallery, 1999).

The time of completion for the re-administration varied from two to four weeks and the response rate was 42.3 %(n=11). One respondent who did not specify his/her name was omitted from the test-retest reliability study. This lowered the response rate to 38.5% (n=10). The intra-class correlation coefficient was 0.65, slightly lower than the commonly accepted level for group comparison.

2.3.6.2 Comparison of responsiveness between the Chinese and English ADDQoL

2.3.6.2.1 Response distribution

Table 2.3 shows the response distribution of unweighted impact scores, importance ratings, weighted impact scores of each domain by two languages. Unweighted impact scores of 4 domains, namely, "family life" [English vs. Chinese: -2.00 (-2.00, 0.00) vs. -2.00 (-2.88, -2.00)], "friendship and social life" [English vs. Chinese: -1.00 (-2.00, 0.00) vs. -2.00 (-2.00, -2.00)], "close personal relationship" [English vs. Chinese: -1.00 (-2.00, 0.00) vs. -2.00 (-2.00, -1.00)] and "dependence on others" [English vs. Chinese: -1.00 (-2.00, 0.00) vs. -2.00 (-2.00, -1.00)] were statistically different ($p<0.05$) between the two language versions with Chinese-speaking respondents reporting lower scores.

Statistically significant difference ($p<0.05$) in importance ratings was only detected in the domain of "employment" with Chinese-speaking respondents reporting higher ratings [English vs. Chinese: 2.00 (0.00, 3.00) vs. 2.00 (2.00, 3.00)]. For weighted impact scores, no statistically significant difference was found between the two languages. In addition, the ranges of responses in terms of "unweighted impact score" and "importance ratings" were quite similar between the two language groups. Interestingly, positive impact of diabetes was reported in the domains of "physical appearance" and "freedom to drink" by one (English-speaking) and three (two English speaking and one Chinese-speaking) respondents respectively.

Table 2.2 Reliability of the Chinese -ADDQOL.

Item	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Squared multiple correlation	Cronbach's alpha if item deleted
Family life	-65.11	1419.23	0.36	0.51	0.94
Friendship and social life	-66.33	1385.02	0.60	0.88	0.94
Close personal relationship	-66.65	1371.92	0.60	0.88	0.94
Sex life	-67.44	1410.69	0.42	0.86	0.94
Physical appearance	-65.85	1346.78	0.66	0.83	0.94
Physical health	-66.52	1381.10	0.73	0.91	0.94
Work (Employment)	-67.10	1368.00	0.57	0.88	0.94
Holiday	-66.73	1369.09	0.71	0.93	0.94
Leisure activities	-66.90	1375.88	0.67	0.96	0.94
Local or long distance journeys	-66.29	1336.44	0.82	0.90	0.94
Self-confidence	-65.54	1286.34	0.90	0.96	0.93
Motivation	-66.12	1337.87	0.75	0.93	0.94
People's reaction	-66.62	1312.41	0.81	0.97	0.94
Feelings about the future	-65.42	1335.29	0.82	0.94	0.94
Financial situation	-65.31	1295.02	0.90	0.97	0.93
Dependence on others	-66.48	1338.85	0.67	0.77	0.94
Living condition	-66.33	1314.90	0.80	0.95	0.94
Freedom to eat	-65.25	1384.91	0.48	0.93	0.94
Freedom to drink	-66.29	1422.88	0.33	0.96	0.95

Table 2.3: Comparison of response distribution between the English and the Chinese sample

Domain	Impact scores unweighted			Importance ratings			Impact scores weighted by importance		
	Median (Interquartile range)			Median (Interquartile range)			Median (Interquartile range)		
	English	Chinese	p-value	English	Chinese	p-value	English	Chinese	p-value
Family life	-2.00 (-2.00, 0.00)	-2.00 (-2.88, -2.00)	0.023*	3.00 (2.00, 3.00)	3.00 (2.00, 3.00)	0.220	-3.50 (-6.00, 0.00)	-6.00 (-6.00, -2.00)	0.185
Friendship and social life	-1.00 (-2.00, 0.00)	-2.00 (-2.00, -2.00)	0.029*	2.00 (2.00, 3.00)	2.00 (1.00, 2.00)	0.063	-2.50 (-6.00, 0.00)	-4.00 (-4.00, -1.75)	0.387
Close personal relationship	-1.00 (-2.00, 0.00)	-2.00 (-2.00, -1.00)	0.047*	2.00 (2.00, 3.00)	2.00 (2.00, 3.00)	0.945	-3.00 (-4.00, 0.00)	-4.00 (-4.50, 0.00)	0.395
Sex life	-1.00 (-2.00, 0.00)	-2.00 (-2.00, -1.00)	0.057	1.50 (0.00, 2.00)	2.00 (1.00, 2.00)	0.426	-0.00 (-4.00, 0.00)	-1.50 (-4.25, 0.00)	0.329
Physical appearance	-1.00 (-2.00, 0.00)	-2.00 (-2.00, -1.00)	0.134	2.00 (1.00, 3.00)	2.00 (1.75, 3.00)	0.830	-2.00 (-4.50, 0.00)	-4.00 (-6.00, -1.00)	0.115
Physical health	-2.00 (-2.00, -0.75)	-2.00 (-2.00, -1.00)	0.531	2.00 (2.00, 2.00)	2.00 (1.75, 2.00)	0.434	-3.50 (-4.00, -0.75)	-4.00 (-4.00, -2.00)	0.538
Work (Employment)	-1.00 (-2.00, 0.00)	-2.00 (-2.00, -0.75)	0.110	2.00 (0.00, 3.00)	2.00 (2.00, 3.00)	0.037*	-1.00 (-4.00, 0.00)	-2.00 (-6.00, 0.00)	0.509
Holiday	-2.00 (-2.00, 0.00)	-2.00 (-2.63, -1.00)	0.280	2.00 (1.00, 2.25)	2.00 (1.75, 3.00)	0.691	-4.00 (-4.00, 0.00)	-3.00 (-6.00, -0.75)	0.790
Leisure activities	-2.00 (-2.00, 0.00)	-1.00 (-2.00, -1.00)	0.951	2.00 (2.00, 2.00)	2.00 (2.00, 2.00)	0.504	-3.50 (-4.00, 0.00)	-2.00 (-4.00, -1.00)	0.979
Local or long distance journeys	-2.00 (-2.00, -1.00)	-2.00 (-2.13, -1.00)	0.631	2.00 (1.00, 2.00)	2.00 (1.00, 2.00)	0.746	-4.00 (-4.00, 0.00)	-4.00 (-5.25, -1.75)	0.454

Table 2.3 (Continued)

Domain	Impact scores unweighted			Importance ratings			Impact scores weighted by importance		
	Mean (SD)			Mean (SD)			Mean (SD)		
	English	Chinese	p-value	English	Chinese	p-value	English	Chinese	p-value
Self-confidence	-2.00 (-2.00, 0.00)	-2.00 (-2.63, -1.00)	0.384	2.00 (2.00, 3.00)	2.00 (2.00, 3.00)	0.868	-4.00 (-6.00, 0.00)	-4.00 (-7.88, -2.00)	0.508
Motivation	-2.00 (-2.00, -1.00)	-2.00 (-2.00, -1.00)	0.206	2.00 (2.00, 2.25)	2.00 (1.00, 3.00)	0.464	-4.00 (-4.00, -1.00)	-4.00 (-6.00, -1.00)	0.353
People's reaction	-1.00 (-2.00, 0.00)	-1.50 (-2.13, -0.75)	0.120	2.00 (1.75, 2.00)	2.00 (1.00, 2.00)	0.483	-2.00 (-4.00, 0.00)	-2.00 (-4.50, -0.75)	0.207
Feelings about the future	-2.00 (-2.00,-1.00)	-2.00 (-2.13, -1.00)	0.584	2.00 (2.00, 3.00)	2.00 (2.00, 3.00)	0.819	-4.00 (-6.00, -1.75)	-4.00 (-6.00, -2.00)	0.533
Financial situation	-2.00 (-2.00, 0.00)	-2.63 (-2.00, -1.00)	0.353	2.00 (2.00, 3.00)	2.00 (2.00, 3.00)	0.475	-4.00 (-6.00, 0.00)	-5.00 (-6.38, -2.00)	0.194
Dependence on others	-1.00 (-2.00, 0.00)	-2.00 (-2.00, -1.00)	0.041*	2.00 (2.00, 3.00)	2.00 (1.00, 3.00)	0.337	-2.00 (-4.00, 0.00)	-3.50 (-6.00, 0.00)	0.147
Living condition	-2.00 (-2.00,-1.00)	-2.00 (-2.00, -0.50)	0.462	2.00 (2.00, 3.00)	2.00 (1.00, 3.00)	0.324	-4.00 (-4.00, -1.00)	-4.00 (-6.00, 0.00)	0.672
Freedom to eat	-2.00 (-3.00, -2.00)	-2.00 (-3.00, -2.00)	0.539	2.00 (2.00, 3.00)	2.00 (1.50, 3.00)	0.792	-4.00 (-9.00, -3.50)	-4.00 (-9.00, -2.00)	0.398
Freedom to drink	-2.00 (-3.00, -1.00)	-2.00 (-2.00, -1.00)	0.885	2.00 (1.00, 3.00)	2.00 (1.00, 2.00)	0.481	-4.00 (-6.00, -1.00)	-4.00 (-4.00, -1.00)	0.436

*: $p<0.05$

The one who reported positive impact of diabetes on "physical appearance" rated "my physical appearance" to be "very important". As for the three respondents who reported positive impact of diabetes on "freedom to drink", two respondents (one English speaking and one Chinese speaking respectively) rated such domain as "somewhat important" while the other English-speaking respondents rated it as "not important at all."

2.3.6.2.2 Use of N/A options

The N/A option was provided for 5 domains, namely, "family life", "close personal relationship", "sex life", "work (employment)" and "holiday". The rank of N/A option use was similar between the two groups with domains of "sex" and "work (employment)" being the top 2. Percentages were also similar in 4 domains except for the domain of "holiday" (English-speaking: 2 (4.8%); Chinese-speaking: 4 (15.4%)) by two languages. (See Table 2.4)

Table 2.4. Use of N/A options between the English and the Chinese samples

	N (%)	
Domain	English	Chinese
Family life	3 (7.1)	3 (11.5)
Close personal relationship	6 (14.3)	4 (15.4)
Sex life	12 (28.6)	7 (26.9)
Work (Employment)	11 (26.2)	8 (30.8)
Holiday	2 (4.8)	4 (15.4)

2.3.6.2.3 Effect of weighting impact ratings

Weighting impact scores by importance ratings apparently changes the ranking of domains, which were originally demonstrated by unweighted scores (see Table 2.5). Except for the two domains of "sex life" and "feeling about future", all the other 17 domains showed slight (1 notch) to distinctive (10 notches) changes in both directions, which indicated the usefulness of weighting impact ratings.

Table 2.5. Comparison of item ranking with and without importance weighting by language

Item	English			Chinese		
	Ranking (Unweighted)	Ranking (Weighted)	Change in order of ranking	Ranking (Unweighted)	Ranking (Weighted)	Change in order of ranking
Family life	9	5	4	2	1	1
Friendship and social life	13	9	4	3	11	-8
Close personal relationship	15	15	0	10	15	-5
Sex life	19	19	0	19	19	0
Physical appearance	14	14	0	13	6	7
Physical health	7	11	-4	12	13	-1
Work (Employment)	18	16	2	16	18	-2
Holiday	8	10	-2	6	16	-10
Leisure activities	10	12	-2	18	17	1
Local or long distance journeys	4	8	-4	7	9	-2
Self-confidence	6	4	2	9	5	4
Motivation	12	13	-1	11	7	4

Table 2.5 (Continued)

Item	English			Chinese		
	Ranking (Unweighted)	Ranking (Weighted)	Change in order of ranking	Ranking (Unweighted)	Ranking (Weighted)	Change in order of ranking
People's reaction	17	18	-1	17	14	-3
Feelings about the future	3	3	0	4	4	0
Financial situation	5	6	-1	8	3	5
Dependence on others	16	17	-1	14	12	2
Living condition	11	7	4	15	10	5
Freedom to eat	1	1	0	1	2	-1
Freedom to drink	2	2	0	5	8	-3

2.3.6.3 Construct validity

All four *a-priori* hypotheses were fulfilled in the study although not all of the results were statistically significant (See Table 2.6).

Convergent and divergent validity was demonstrated by the following correlations: "Present HRQoL" scores correlated moderately with EQ-5D utility scores (r=0.268, p=0.185), and SF-6D index scores (r=0.351, p=0.078). As predicted, ADDQoL median weighted scores correlated moderately with "HRQoL without diabetes" scores (r=0.339, p=0.090) but weakly with "Present HRQoL" scores (r=0.027, p=0.896).

As for the known group validity, the results showed that lower ADDQoL median weighted scores moderately correlated with lower SF-6D mental scores (r=0.247, p=0.224), suggesting that those who were more depressed had a poorer HRQoL; and FFM scores positively correlated with ADDQoL scores (r=0.288, p=0.182), which provided the finding of "better HRQoL is associated with better family functioning" among the diabetes patients.

Table 2.6: Spearman rank correlation among Chinese-ADDQOL weighted scores," Present HRQoL", "HRQoL without diabetes", EQ-5D utility, SF-6D and FFM scores

	ADDQOL scores			EQ-5D utility scores	SF-6D		FFM
	"Present HRQoL"	Weighted	"HRQoL without diabetes"		Index	Mental	
ADDQoL							
Present HRQoL scores	1.000						
Weighted scores	0.027	1.000					
HRQoL without diabetes scores	-0.015	0.339	1.000				
EQ-5D utility scores	0.268	0.238	0.021	1.000			
SF-6D							
Index score	0.351	0.256	-0.016	0.785**	1.000		
Mental score	0.287	0.247	-0.206	0.342	0.597	1.000	
FFM score	-0.121	0.288	0.271	0.362	0.378	-0.148	1.000

*p<0.05, **p<0.01, ***p<0.001

2.3.7 Functional equivalence

The results of the conceptual equivalence, item equivalence, semantic equivalence, operational equivalence and measurement equivalence listed above indicated that a reasonable degree of equivalence was achieved in all the 5 areas. Therefore, the functional equivalence between the Chinese-ADDQoL and English-ADDQoL was demonstrated in this study.

2.4 Discussion

Due to increasingly higher prevalence and huge economic burden of diabetes among Chinese-speaking population, it becomes important and urgent to implement cost-effective diabetes management programs to ensure the sustainability of healthcare and economic resources. Hence, a well translated and culturally adapted disease-specific HRQoL measure such as ADDQoL could contribute to the more accurate assessment of the effectiveness of the disease management programs, compared with generic HRQoL measures.

To the best of our knowledge, the current study is the first to translate and culturally adapt the original English-ADDQoL into the Chinese version. The promising findings of the Singapore-based study could contribute to further adaptation of this Chinese-ADDQoL among other Chinese-speaking populations in the world.

Furthermore, the current study also serves as a pioneering yet important example to apply the systematic universalist approach to translate and adapt an English HRQoL instrument into a Chinese version to ensure equivalences in the full spectrum. Unlike some other equivalent studies, the current study emphasized the importance of conceptual equivalence and item equivalence as the prerequisite for the translation of ADDQoL, which is exceptionally important for the cross-cultural adaptation. The translation process was integrated into the whole adaptation process rather than simply heavily relying on translation techniques without cognitive debriefing. Therefore, the mistake of rigid assumption that HRQoL instruments would be equally valid in any culture was avoided. We thus carried out the investigation of measurement equivalence only when conceptual equivalence, item equivalence, semantic equivalence and operational equivalence had been demonstrated.

The main findings suggest that the Chinese-ADDQoL adapted in Singapore achieved all six types of equivalences with the original English-version developed in UK. In the mean time, it is worthwhile pointing out the following findings or issues, as they could provide either precautions or suggestions for future studies.

First, in this study, the test-retest reliability was relatively poor with its ICC (0.65) a little below the acceptability level of 0.7 (Nunnally, 1997). The relatively poorer test-retest reliability might be due to insufficient sample size (n=10) for data analysis and poor compliance of the second administration. Of the 10 respondents of test-retest reliability, 3 filled up the retest questionnaires at around the 4^{th} week after the first administration, which violated the preset 2-week interval rule. It was also likely that some of their answers were different due to recall problem or certain change in the disease state, since the 4-week interval might be too long. Therefore, further test-retest reliability studies on Chinese-ADDQoL should be conducted for verification of the results. Moreover, methods to increase response rates to postal questionnaires as suggested by Edwards et al. (2007) could be used for retests, such as pre-notification, giving out small incentives, etc.

Second, we have observed interesting findings regarding the differences in response distributions as reported by the English-speaking and Chinese-speaking study subjects. It was shown that the unweighted impacts scores of four domains, namely, "family life", "friendship and social life", "close personal relationship" and "dependence on others" were rated as significantly worse by the Chinese-speaking subjects than their English-speaking counterparts. Such differences might be related to the differences in the subject characteristics as we have found, in which Chinese-speaking subjects had a relatively poorer socioeconomic status especially in the areas of family functioning, education levels, employment status, and dwelling type. Similarly, the domain of "employment" was again found to be significantly more important as rated by Chinese-speaking subjects as compared to the English-speaking counterparts who had a much higher employment ratio. Yet when the comparisons were made by weighted ADDQoL scores, no statistically significant differences were detected between the two study groups. We suggest that when the margin for the minimal clinically important difference is established for ADDQoL in other clinical studies, further analysis could be performed to examine whether the score differences

after adjustment for the confounding variables in baseline characteristics between the two study groups are clinically unimportant so as to demonstrate the measurement equivalence in the aspect of responsiveness (Drasgow and Kanfer, 1985).

Third, the use of N/A options and weighted scoring method were proven to be necessary and effective for Chinese-ADDQoL as well as English-ADDQoL. Such results further emphasized the importance of increasing relevance and sensitivity of a HRQoL questionnaire for better accuracy.

Last but not least, due to the nature as an exploratory study, we only recruited a small sample size as a start-off to explore the translation and adaptation process by the universalist approach. Hence, the small sample size would hamper generalizing our results to the general Chinese-speaking diabetic population in Singapore. Nevertheless, the preliminary study did provide conceptual background and encouraging results for using the universalist approach in the translation and cultural adaptation of HRQoL instrument from West to East in an effective and efficient manner.

2.5 Conclusion

The results of this exploratory study suggest that this Singaporean Chinese-ADDQoL has achieved functional equivalence with both the original and the Singaporean adapted English-ADDQoL by demonstrating conceptual equivalence, item equivalence, semantic equivalence, operational equivalence and measurement equivalence. It also shows the strength of using the universalist approach to ensure high-quality translation and cultural adaptation process in future studies.

Chapter Three

A Qualitative Review of Factors Influencing Cost-Utility Analysis in assessing disease management

3.1 Introduction

Despite the fact that health economic evaluation is a relatively new research area in Asia, its importance as a quantitative tool to assist decision-making in disease management has been increasingly recognized due to the escalating healthcare expenditures and tightening budget constraints. As mentioned in Chapter 1, cost-utility analysis (CUA) is a special type of economic evaluation that measures health benefits in utility-weighted life years and computes a cost per utility-measure ratio, which has the crucial advantage over others to enable comparisons across different disease management programs (Gerard, 1992). Therefore, CUA has been recommended as the preferred form of reporting health economic evaluation results by the Panel on Cost-Effectiveness in Health and Medicine (Siegel et al., 1996).

However, surprisingly, even in Western countries, where CUA has been widely applied in evaluating disease management programs, its impact in real-world medical decision-making has been shown to be rather limited across various countries. According to a focus group research study among a panel of decision-makers from two UK health authorities, one of the key problems was the lack of generalizability or transferability of the results published in most of the medical literature (Hoffmann et al., 2002). Similarly, decisions-makers and researchers from other Western countries like Canada (PausJenssen et al., 2003), Australia (Salkeld et al., 1995), Sweden (Anell and Persson, 2005) have also identified such problems as a crucial barrier of using CUA as a valid and reliable tool in their decision-making processes. This lack of generalizability threatens the relevance of CUA, which has also been commented by researchers from Asian countries like Lee et al. (2005, Korea) and Ikegami et al. (2002, Japan). These findings challenged the opinion of CUA as a useful and accurate

decision informing tool that could be further widely applied in Asian countries. Hence, it is of paramount importance to explore what are the factors have jeopardized the generalizability of CUA results.

Actually, when trying to explore the reasons causing the perceived lack of generalizability among end-users of CUA or other economic evaluations in general, it becomes obvious that the uncertainties relating to the impact on outcomes caused by the various factors tested in CUA is a major hindrance. Hence, one major hurdle affecting the generalizability of CUA results is, therefore, the lack of consensus about factors that should be tested when examining the applicability in different settings of the results from a CUA.

Naturally, there are many factors and variables that could potentially affect the outcomes when conducting CUA to assess the cost-effectiveness of interventions or disease management programs; and their impact on outcomes are usually tested in sensitivity analyses. Although sensitivity analyses have been widely performed to examine uncertainties caused by the various factors in CUA, it was found that the set of parameters tested in sensitivity analysis was often limited and the overall reporting quality was far from satisfactory (Briggs and Sculpher,1995; Schakman et al., 2004).

In fact, Walker and Fox-Rushby (2001) commented that the very first step to design a well-justified sensitivity analysis was to identify all the potential parameters that are relevant to the model. Based on such list, researchers could then choose the parameters with uncertainties relevant to the case being evaluated to perform sensitivity analysis; and decision-makers can use the check list to assess the

generalizability of the results from the CUA and its sensitivity analyses. Such comment is deemed as valid for the CUA studies universally.

However, an extensive literature search revealed no published literature synthesis focusing on the factors influencing CUA. Hence, to fill in such knowledge gap, the current study was aimed to provide a qualitative literature review of all the influential factors identified in the published CUA studies, regardless of its origin.

3.2 Methods

3.2.1 Literature search strategy

Potentially relevant articles were identified by electronic database search as well as manual search. At first, search of several electronic databases, namely, Medline (1966 to May 2005), PsycINFO (1967 to May 2005), ISI Web of Knowledge (ISI) [including Web of Science (1980 to May 2005), Current Contents Connect (1998 to May 2005), ISI proceedings (2002 to May 2005)], EconLit (1969 to May 2005) and the Cochrane Library (1996-May 2005) was performed, where cost-utility analysis was entered as a free term in the text word search. At this stage, non-English publications whose abstracts were in English were also included as they were more likely to explore factors pertaining to different socio-cultural contexts. It was agreed by the candidate and the two supervisors that a brief summary of non-English publications would be given in this literature review for completeness. However, due to a lack of language proficiency other than English on the part of the candidate, non-English articles were excluded from further full-text review.

Manual search of several key health economics journals including Value in

Health (May 1998-May 2005), Pharmacoeconomics (Jan 1992-May 2005), Health Economics (April 1992-May 2005) and Journal of Health Economics (1982-May 2005) was also performed to identify any articles that were missed out in the electronic database search. As cost-utility analysis has also been referred to as cost-effectiveness analysis or economic evaluation in titles of publications, all of the three terms were used as key words in the manual search. In addition, bibliographies of all articles selected for full-text review were also examined and potentially relevant articles were further selected for full-text review.

3.2.2 Selection of articles for full-text review

All of the identified abstracts retrieved from each of the electronic databases were first downloaded into Reference Manager Version 11 and checked for duplication. All abstracts from the manual search and the modified Reference Manager Database were then examined by the candidate based on the predefined inclusion criteria to select articles for full-text review. Any uncertainties were brought up for discussion with the two supervisors and then finally resolved by reaching consensus after the discussion.

Inclusion criteria were as follows: (1) for empirical studies, factors affecting cost-utility ratio should be clearly identified and quantitatively analyzed by sensitivity analysis; (2) for review articles, factors affecting cost-utility ratio should be clearly qualitatively analyzed in terms of why or how they affected results of CUA; (Denzin and Lincoln, 2000) (3) articles for full-text review should be written in English.

3.3 Results

3.3.1 A brief summary of non-English publications

Fifteen non-English languages articles (4 in Japanese; 5 in Dutch; 2 in Norwegian; 1 each in Korean; German; Spanish and French respectively) were identified as articles dedicated to study of CUA. Based on their English-written abstracts, four articles mentioned about the factors or issues influencing CUA. These were the articles by De Neeling (2004), Belouet et al. (1999), Kristiansen et al. (1997), and Rubeo-Terres et al. (2003) in Dutch, French, Norwegian and Spanish respectively. The factors affecting CUA identified by these articles are summarized in the next paragraph.

De Neeling (2004) argued that quite a few uncertainties restricted the application of CUA for direct comparisons of efficiency across all types of healthcare services. These uncertainties included QALY elicitation methods, QALY elicitation population (patients or the general public), types of costs to be incorporated, discounting of future cost and benefits, etc. In the other three abstracts, (Belouet et al., 1999; Kristiansen et al., 1997; and Rubeo-Terres et al., 2003], disease progression rate, risk of disease worsening or complications, cost of intervention, duration of treatment and discount rate were identified to affect cost-utility ratio in sensitivity analyses, but magnitude of impacts were not mentioned in detail.

3.3.2 Identification of potentially relevant articles for full-text review

Figure 3.1 presents the article selection process together with the results. Before checking for duplication, 2,644 articles were identified from electronic database search (660 from Medline, 36 from Psycinfo, 734 from ISI, 19 from EcoLit and 1,195 from Cochrane Library). After checking for duplication using Reference Manager, 716 articles were identified without duplication. After reviewing the titles and

abstracts according to the predefined criteria, 93 articles were selected for full-text review. The other 622 articles were not included due to the following reasons: (1) it was not a CUA study or a CUA-related review (N= 275); (2) there was no sensitivity analysis reported (N=145); (3) there was no influential factor identified from sensitivity analysis (N=202).

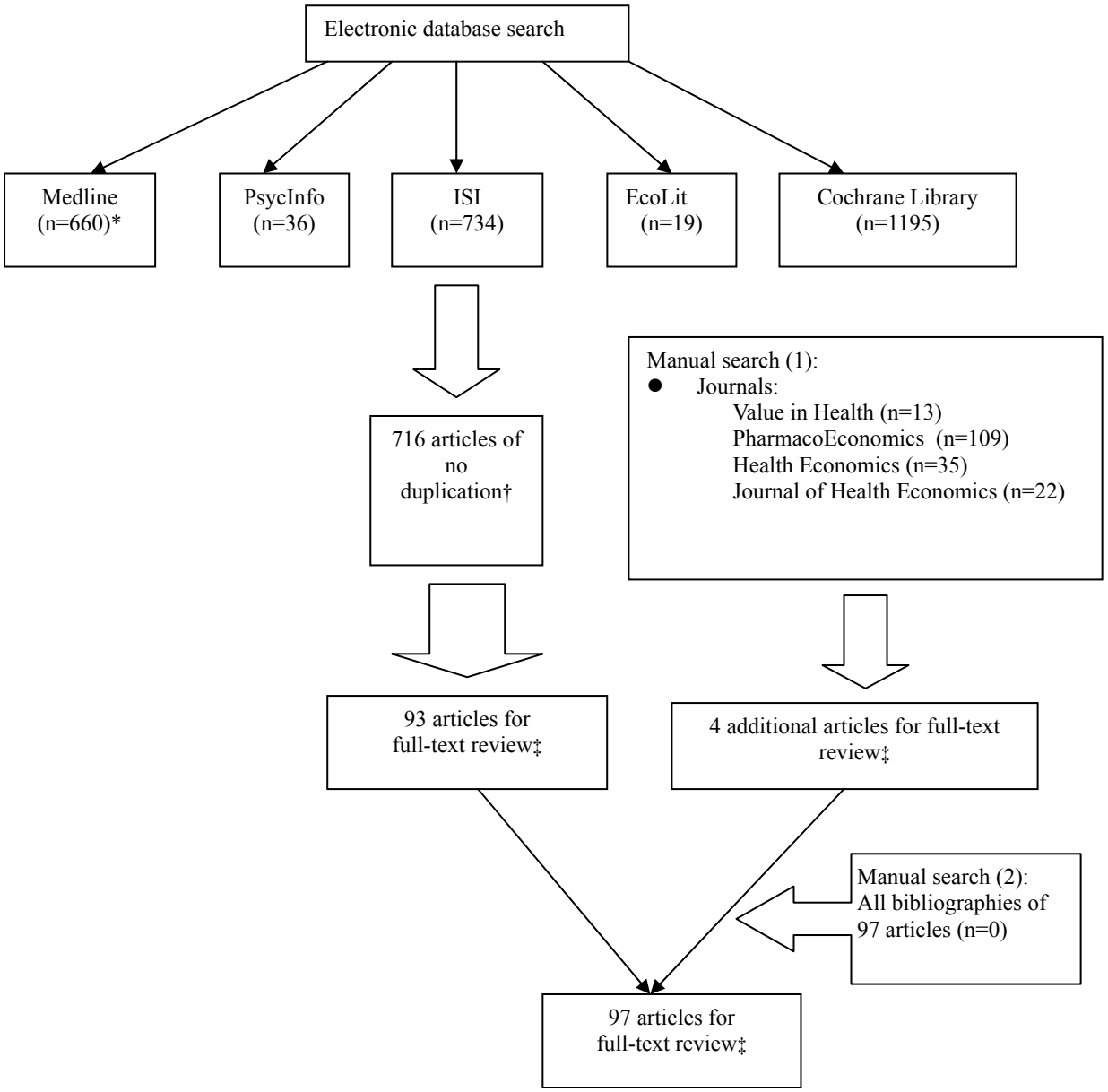
As for the manual search, four additional articles were identified for full-text review. Hence, bibliographies of 97 articles were examined for full-text review, but no additional article was further identified which had not been identified through the electronic literature search.

Of the 97 articles included for full-text review, 5 were review articles and 92 were empirical studies. The majority was conducted in the United States of America (N=41), United Kingdom (N=28), and Canada (N=14). With the exception of one study performed in Hong Kong, the rest of the studies were carried out in other Western countries including the Netherlands (N=5), Germany (N=2), Spain (N=2) Australia, Switzerland, Belgium, Denmark and Mexico (N=1 for each country).

3.3.3 Factors affecting cost-utility ratio by categories

Altogether 20 factors that could potentially impact on the outcomes of a CUA were identified in this literature review. These factors could be categorized into several broad groupings as treatment-related, disease-related, patient-related, cost-related, and methodology-related.

Figure 3.1 Article selection process



*: Numbers refer to references identified
†: Duplication was checked by Reference Manager.
‡: Inclusion criteria were as follows: (1) for empirical studies, factors affecting cost-utility ratio should be clearly identified and quantitatively analyzed by sensitivity analysis; (2) for review articles, factors affecting cost-utility ratio should be clearly qualitatively analyzed in terms of why or how they affected results of CUA; (3) articles for full-text review should be written in English;

3.3.3.1 Treatment-related factors

Efficacy (N=12), duration (N=4), and frequency of treatment (N=1) were the three treatment-related factors that have been found to directly influence cost-utility ratio in a wide range of diseases such as myocardial infarction (Sanders et al., 2001), diabetes (Kiberd and Larson, 2000), breast cancer (Li et al., 2001), HIV (Pinkerton et al., 2000), etc. As shown in Table 3.1, duration and efficacy of treatment are the two factors that had negative impacts on cost-utility ratios, while the impact of frequency of treatment was reported to be positive (where negative impact denotes an increase in the value of factor led to a decrease in cost-utility ratio; while positive impact denotes an increase in the value of factor led to an increase in cost-utility ratio.)

Ten out of 12 studies reported that changes in the efficacy of treatment exerted minor and negative impacts on cost-utility ratios, but the impacts were not large enough to cross the preset thresholds (USD 20,000/QALY, or USD 50,000/QALY, or USD 100,000/QALY depending on individual cases) to alter the conclusion about cost-effectiveness of the interventions. On the contrary, increases in duration of treatment were mostly found to have major and negative impacts on cost-utility ratios (three out of four studies). For example, in a CUA of finasteride for treatment of benign prostatic hyperplasia, it was found that duration of treatment was the critical factor in judging cost-effectiveness of the drug. Finasteride was shown to be the dominant alternative compared with both transurethral resection of the prostate (TURP) and watchful waiting for patients with moderate symptoms, only when the duration of drug therapy was three years or less (Baladi et al., 1996).

Comparatively, although frequency of treatment was identified to be the

influential factor in only one comparative study assessing the cost-effectiveness of sildenafil versus papaverine-phentolamine injections for treatment of erectile dysfunction (Schleinitz and Heidenreich, 2005), the impact turned out to be quite critical. Doubling the frequency of use of sildenafil was shown to almost increase cost per QALY by two-fold. It was mainly because that doubling the dosing frequency doubled the drug acquisition cost, which was the major cost driver of the sum of cost in the study. However, comparatively, the effect on QALY turned out to be a just slight increase, which was far from being doubled.

Table 3.1. Treatment-related factors that affect cost-utility ratio

Factor	Disease	Reference	Magnitude of impact on cost–utility ratio*	Hypothesis or justification provided by author(s), if any
Duration of treatment				
	Chronic myeloid leukemia	Gordois et al., 2003	-	Uncertainty regarding the duration of efficacy
	HIV	Pinkerton et al., 2000	--	Uncertainty regarding the duration of efficacy
	Benign prostatic hyperplasia	Baladi et al., 1996	--	Uncertainty regarding the duration of treatment
	Dystonia	Gudex et al., 1997	--	Uncertainty about the duration of treatment
Efficacy of treatment				
	Myocardial infarction	Sanders et al., 2001;	--	Uncertainty about the efficacy
	Breast cancer	Hayman et al., 2000;	-	Uncertainty about the efficacy
		Hutton et al., 1996	-	Uncertainty about the efficacy
		Verma and Rocch, 2003i	-	Uncertainty about the efficacy
	Diabetes	Czoski-Murray et al., 2004	-	Uncertainty about the efficacy
	Prostate cancer	Krahn et al., 1994	-	Uncertainty about the efficacy
	Parkinson's disease	Nuijten and Rutten, 2003	N.A.	Efficacy of treatment could be a potential source of uncertainty, so it should be incorporated into sensitivity analysis

Table 3.1 (Continued)

Factor	Disease	Reference	Magnitude of impact on cost–utility ratio*	Hypothesis or justification provided by author(s), if any
Efficacy of treatment				
	Back pain	Rivero-Arias et al., 2005	-	Uncertainty about the efficacy
	coronary syndrome	Schleinitz & Heidenreich, 2005	-	Uncertainty about the efficacy
	Hepatitis C	Stein et al., 2004	-	Uncertainty about the efficacy
	Orbital foreign body	Seidenwurum et al., 2000	-	Uncertainty about the efficacy
	chronic obstructive pulmonary disease	Añón et al., 1999	-	Uncertainty about the efficacy
Frequency of treatment	Erectile dysfunction	Stolk et al., 2000	++	Uncertainty about the frequency of use

*:
+: minor positive influence, indicating that increase in the factor would also result in minor increase in the cost-utility ratio but the impact was not large enough to change the conclusion regarding cost-effectiveness of the intervention ;
++: major positive influence, indicating that increase in the factor would also result in a substantial increase in the cost-utility ratio and the impact was large enough to change the conclusion regarding cost-effectiveness of the intervention;
−: minor negative influence, indicating that increase in the factor would result in minor decrease in cost-utility ratio but the impact was not large enough to change the conclusion regarding cost-effectiveness of the intervention ;
−−: major negative influence, indicating that increase in the factor would result in a substantial decrease in cost-utility ratio and the impact of the factor was large enough to change the conclusion regarding cost-effectiveness of the intervention;
N.A.: Not available (magnitude of the impact was not presented in the publication).
Unless indicated in the study, the threshold of cost-effectiveness was USD 50,000/QALY in this review.

3.3.3.2 Disease-related factors

There were altogether six disease-related factors identified as shown in Table 3.2, namely, risk level [N=18: including baseline risk of disease (N=7); and risk of disease worsening or complication or even death (N=11)], prevalence rate (N=3), severity level (N=5), survival length (N=3), disease progression rate (N=2) and incidence rate (N=1). Interestingly, these six factors could be further related to the type of intervention, that is, treatment or prevention.

For treatment programs, cost-utility ratios were shown to be affected by risk level (disease worsening or complication or even death), severity level, disease progression rate and survival length.

Among these four factors, risk level and severity level were reported to have minor or major and negative impacts on cost-utility ratios (Table 3.2). For example, a study on the cost-effectiveness of prophylactic use of the implantable cardioverter defibrillator (ICD) or amiodarone after myocardial infarction showed that risk of ejection fractions played a key role in determining cost-effectiveness of the treatment (Sanders et al., 2001). For moderate efficacies, when the risk level was less than or equal to 0.3, 0.31 to 0.4, or greater than 0.4, the cost-effectiveness of amiodarone compared with no therapy was $43,100/QALY, $66,500/QALY, and $132,500/QALY, respectively. Under the same condition, the cost-effectiveness of ICD compared with amiodarone was $71,800/QALY, $195,700/QALY, and $557,900/QALY. As for severity level, in a CUA study conducted by Johnston et al. (1999), it was found that treatment of small, asymptomatic, unruptured cerebral aneurysms worsened clinical

outcomes, and thus was neither effective nor cost-effective, compared with the same treatment on patients with aneurysms that were ≥10 mm or symptomatic.

Comparatively, the other two factors, namely, disease progression rate and survival length were found to have minor and positive influences. It was found that reduced disease progression rate increased QALYs, but did not substantially reduce direct medical costs (Danese et al., 1996). As for survival length, its positive impact might be more related with the increase of total medical cost due to the extension of survival length, while the improvement in QALY was not comparatively significant (Richards and Irving, 1996; Manns et al., 2003; Bakhai et al., 2003).

For disease prevention programs, cost-utility ratio was found to be influenced by baseline risk of disease, prevalence rate and incidence rate. All of the three factors had negative impacts on cost-utility ratios, the magnitude of which ranged from minor to major (Table 3.2). For example, in the CUA of a HIV prevention intervention for mentally ill adults, compared with other health promotion interventions, it was marginally cost-effective if the analysis was restricted to the subset of high-risk and sexually-active women (Pinkerton et al., 2001). In another study of the HIV risk reduction intervention among African-American male adolescents, the HIV prevalence rate in the community had a substantial impact on the cost-utility ratio. The program would no longer be cost-effective if the prevalence rate was less than 0.58% for all youth despite past sexual experience or less than 0.35% for sexually active youth (Pinkerton et al., 2000). Similar impact was also indicated by incidence rate, which was identified to be the most influential parameter in the sensitivity

analyses of a screening test for hepatocellular carcinoma in patients with hepatitis C cirrhosis (Arguedas et al., 2003).

Table3.2. Disease-related factors that affect cost-utility ratio

Factor	Disease	Reference	Magnitude of impact on cost–utility ratio*	Hypothesis or justification provided by author(s), if any
Severity level				
	Vestibular aqueduct syndrome	Bichey et al., 2002	-	Uncertainty about severity level
	Severe proliferative vitreoretinopathy	Brown et al., 2002	-	Uncertainty about severity level
	Pediatric cochlear implantation	Bichey et al., 2002	-	Uncertainty about severity level
	Parkinson's disease	Nuijten and Rutten, 2003	N.A.	Severity level could be a potential confounding variable of the effectiveness of treatment, so it should be incorporated into sensitivity analysis
	Cerebral aneurysms	Johnston et al., 1999	--	Hypothesized that patients of different severity would lead to difference in cost-effectiveness of the treatment.
Risk level				
(1) Disease worsening or complications or death				
	Renal transplantation	Keown et al., 2004	--	Uncertainty about the risk of rejection after renal transplantation; the higher the rejection possibility, the lower the cost-utility ratio
		Keown et al., 2004, 2001	--	Uncertainty about the risk of rejection after renal transplantation
	Abdominal aortic aneurysm (AAA)	St Legar et al., 1996	-	Uncertainty about the annual risk of rupture for AAA

Table3.2 (Continued)

Factor	Disease	Reference	Magnitude of impact on cost–utility ratio*	Hypothesis or justification provided by author(s), if any
(1) Disease worsening or complications or death				
	Diabetes	Vijan et al., 2000	--	The risk reduction varies dramatically by age and level of glycemic control.
	Myocardial infarction	Sanders et al., 2001	--	Uncertainty about the risk of death
	Allogeneic transfusion	Sonnernburg et al., 1999	--	Uncertainty about the relative risk of bacterial infection following allogeneic transfusion
	Appendectomy wounds	Brasel et al., 1997	--	Uncertainty about the risk of primary infection
	Rheumatic diseases	Gabriel et al., 1995	--	Uncertainty about the risk of disease
	Chronic arthritis	Zabinski et al., 2001	-	Uncertainty about the GI risk in the population
	Pancreatitis,	Gregor et al., 1996	--	Uncertainty about having an occult common duct stone
(2) Baseline risk of disease				
	Acute coronary syndrome	Latour-Pérez et al., 2004	-	Uncertainty about baseline risk of cardiovascular events
	Meningeal signs	Oosteinbrink et al., 2002	--	Uncertainty about the risk of bacterial meningitis or sequelae
	HIV	Pinkerton et al., 2000	--	Hypothesized that if the analysis was restricted to those youth who reported sexual activity in the 3 months preceding the intervention
		Pinkerton et al.,1997	-	Uncertainty about the per-contact risk of HIV transmission
		Pinkerton et al., 2001	-	Uncertainty about risk of HIV

Table3.2 (Continued)

Factor	Disease	Reference	Magnitude of impact on cost–utility ratio*	Hypothesis or justification provided by author(s), if any
(2) Baseline risk of disease				
	Chronic arthritis	Spiegel et al., 2003	--	Uncertainty about the baseline risk as a having a history of bleeding ulcers
	Diabetes	Ragnarson and Apelgvist, 2001	--	Uncertainty about the risk of foot ulcers and lower extremity amputations
Incidence rate				
	Hepatitis C	Arquedas et al., 2003	--	Uncertainty about the incidence rate of hepatitis C
Disease progression rate				
	Parkinson's disease	Nuijten and Rutten, 2003	N.A.	Disease progression rate could be a potential confounding variable of the effectiveness of treatment, so it should be incorporated into sensitivity analysis
	Diabetes	Kiberd and Larson	+	Uncertainty about the progression rate to the end stage renal disease
	Mild thyroid failure	Danese et al., 1996	+	Uncertainty about the disease progression rate
Survival length				
	Coronary heart disease	Chau et al., 2001	+	Uncertainty about the survival length
	End-stage renal disease	Kiberd and Larson, 2000	+	Uncertainty about the survival length
	Intestinal failure	Richards and Irving, 1996	+	Uncertainty about the survival length
Prevalence rate				
	Orbital foreign body	Seidenwurum et al., 2000	-	Uncertainty about the prevalence rate
	Orthoptic screening	König and Barry, 2004	-	Uncertainty about the prevalence rate

Table3.2 (Continued)

Factor	Disease	Reference	Magnitude of impact on cost–utility ratio*	Hypothesis or justification provided by author(s), if any
Prevalence rate				
	HIV	Pinkerton et al., 2000	--	Uncertainty about the HIV prevalence in the community

*:

+: minor positive influence, indicating that increase in the factor would also result in minor increase in the cost-utility ratio but the impact was not large enough to change the conclusion regarding cost-effectiveness of the intervention ;

++: major positive influence, indicating that increase in the factor would also result in a substantial increase in the cost-utility ratio and the impact was large enough to change the conclusion regarding cost-effectiveness of the intervention;

−: minor negative influence, indicating that increase in the factor would result in minor decrease in cost-utility ratio but the impact was not large enough to change the conclusion regarding cost-effectiveness of the intervention ;

−−: major negative influence, indicating that increase in the factor would result in a substantial decrease in cost-utility ratio and the impact of the factor was large enough to change the conclusion regarding cost-effectiveness of the intervention;

N.A.: Not available (magnitude of the impact was not presented in the publication).

Unless indicated in the study, the threshold of cost-effectiveness was USD 50,000/QALY in this review.

3.3.3.3 Patient-related factors

Incorporating patient-specific characteristics or performing subgroup analysis in CUA studies have been increasingly advocated, because such information would improve the validity and comparability of results across different studies or populations (Van Hout et al., 1999; Arneson and Trommald, 2005; Gudex et al., 1997). Age (N=6), gender (N=2), compliance/adherence (N=2), race/ethnicity (N=1) were identified as patient-related influential factors as shown in Table 3.3.

According to a literature review by Russel and Sisk (2000), they found that age had been incorporated into several CUA due to its profound impact on life expectancy, risk level, incidence and prevalence rates, and treatment efficacy, which would ultimately influence both cost and utility components of CUA. It was interesting to note that the direction and magnitude of influence was not congruent in the findings, which was dependent on how age affected other QALY- and cost-related variables. For example, when the incidence rate declined with age, the pneumococcal conjugate heptavalent vaccine became far less cost-effective when covering all children less than 60-months of age, compared with those less than 24-months of age (Schleinitz and Heidenreich, 2005). On the contrary, when incidence rate of mild thyroid failure was expected to increase with age, the cost-effectiveness became more favorable when age at first screening was increased (Danese et al., 1996).

However, it was also noted that many studies had not modeled age differences in sufficient detail to ensure that differences in cost-utility ratios by age were accurate. As the influence of age on CUA would also be dependent on both the disease type

(chronic or acute) and treatment type (preventive, curative or palliative), it was suggested that more complete age-specific data should be incorporated to ensure a sound basis for decisions (Richards and Irving, 1996).

Apart from age, gender and race/ethnicity were the other two demographic parameters shown to impact on cost-utility ratios of patients with heart diseases (Kupersmith et al., 1995) and HIV (Johnson-Masotti et al., 2000). Again, there was no definitive conclusion about the direction of impact on cost-utility ratio by these two factors, as it would also be potentially relied on how gender or ethnicity affected other disease-, treatment- or even cost-related variables. Although not explicitly studied in CUA, some of other sociodemographic parameters (e.g. education level, marital status, family support, etc) might indirectly impact on cost-utility ratio due to their influence on health preferences elicited from patients (Goold and Vijan, 1998).

Compliance with treatment was identified to be the other critical parameter in the cost-effectiveness of interventions. It was shown to be the most sensitive parameter in the CUA on patients with depression (Revicki et al., 1997) and diabetes (Clark et al., 2000). For instance, a CUA of an Angiotensin-Converting Enzyme (ACE) inhibitor therapy for diabetic nephropathy, changes in the compliance rate from 67% to 51% could result in a swing from a savings of $899 to an expenditure of more than $1 million per additional QALY (Van Hout et al., 1999).

Table 3.3 Patient-related factors that affect cost-utility ratio

Factor	Disease	Reference	Magnitude of impact on cost–utility ratio*	Hypothesis or justification provided by author(s), if any
Age				
	Intestinal failure	Richards and Irving, 1996	++	Subgroup analysis for different age groups
	Diabetes	Goold and Vijan, 1998	N.A.	It is likely the cost and benefits will be based on the age of patient.
	Non-ST-segment elevation acute coronary syndrome	Latour-Pérez, 2004	++	Uncertainty about the age effect.
	Pneumococcal conjugate immunization	Ess et al., 2003	++	Uncertainty about the age effect
	Mild thyroid failure	Danese et al., 1996	-	Uncertainty about the age effect ; Potential correlation between age and incidence rate
	N/A (review)	Russel and Sisk, 2000	N.A.	It is not known how well studies model the risks and costs associated with age.
Gender				
	HIV	Johnson-Masotti et al., 2000	Major impact (all three interventions were cost-effective for men, but only single-session was cost-effective for woman)	Uncertainty about the gender impact
	Mild thyroid failure	Danese et al., 1996	Moderate impact (more favorable towards woman)	Uncertainty about the gender effect ; Potential correlation between gender and incidence rate
Race/ethnicity				
	Diabetes	Goold and Vijan, 1998	N.A.	Potential impact of ethnicity gap on the cost and effectiveness of treatment
Compliance				
	Diabetic nephropathy	Clark et al., 2000	--	Uncertainty about the compliance rate

Table3.3 (Continued)

Factor	Disease	Reference	Magnitude of impact on cost–utility ratio*	Hypothesis or justification provided by author(s), if any
Compliance				
	Depression	Revicki et al., 1997	-	Uncertainty about the compliance rate

*:
+: minor positive influence, indicating that increase in the factor would also result in minor increase in the cost-utility ratio but the impact was not large enough to change the conclusion regarding cost-effectiveness of the intervention ;
++: major positive influence, indicating that increase in the factor would also result in a substantial increase in the cost-utility ratio and the impact was large enough to change the conclusion regarding cost-effectiveness of the intervention;
-: minor negative influence, indicating that increase in the factor would result in minor decrease in cost-utility ratio but the impact was not large enough to change the conclusion regarding cost-effectiveness of the intervention ;
--: major negative influence, indicating that increase in the factor would result in a substantial decrease in cost-utility ratio and the impact of the factor was large enough to change the conclusion regarding cost-effectiveness of the intervention;
N.A.: Not available (magnitude of the impact was not presented in the publication).
Unless indicated in the study, the threshold of cost-effectiveness was USD 50,000/QALY in this review.

3.3.3.4 Cost-related factors

Treatment cost (N=17), incorporation of indirect cost (N=3) were identified to be the two cost-related factors influencing outcomes of CUA as shown in Table 3.4.

Treatment cost was found to be the most frequently identified cost-related parameter in this review. Variations in treatment cost were assumed in different scenarios, such as mere increase or decrease in price/cost of the same drug/device, incorporation of other drugs/devices to treat side effects or complications and changes in the frequencies/doses of the intervention (Van Hout et al., 1999; Earle et al., 2004; Dranitsaris and Hsu, 1999; Stolk et al., 2000).

In some cases, different types of changes in cost were studied both independently and in combination. For example, in a CUA of second-line irinotecan for metastatic colon carcinoma in a multi-center, open-label phase III clinical trial (Earle et al., 2004), it was found that cost-utility ratio was very sensitive to the cost of irinotecan with a base-case ratio of $78627/QALY. If cost of irinotecan was decreased to around 80% of the current cost, then the every-3-week regimen became more attractive than the weekly regimen. Hospitalization, mostly due to admissions for toxicity, was more likely in the weekly regimen. If such excess of hospital costs were removed, then the cost-utility ratio would rise to $108,438/ QALY. A multivariate sensitivity analysis eliminating all other cost offsets indicated a maximum cost-utility ratio of $267,287/QALY. Fifty-five percent of patients in the every-3-week arm started irinotecan at 300 mg/m^2. However, if the intended starting dose were reduced to 300 mg/m^2 among all patients while achieving the same outcomes, then the cost-utility

ratio would decrease slightly to $72,761/QALY.

Recent studies based on lifetime utility maximization models suggest that CUA should account for all future costs, not only direct and non-direct medical cost, but also indirect cost. Especially for CUA from a societal perspective, whether indirect cost is incorporated is a major concern (Liljas, 1998). The magnitude of indirect cost has been shown to substantially alter both the absolute and relative cost-effectiveness of medical interventions, particularly when an intervention would increase length of life more than quality of life. For example, the inclusion of indirect cost significantly reduced the incremental cost-utility ratio. This further demonstrated the cost-effectiveness of intensive medical interventions that decreased mortality among young adults with type 1 diabetes (Meltzer, 1996; Meltzer et al., 2000).

Table 3.4 Cost-related factors that affect cost-utility ratio

Factor	Disease	Reference	Magnitude of impact on cost–utility ratio*	Hypothesis or justification provided by author(s), if any
Treatment cost				
	Diabetes	Kiberd and Larson, 2000	++	Uncertainty about the pancreas transplantation cost
	Orbital foreign body	Seidenwurm et al., 2000	++	Uncertainty about the screening cost
	Chronic arthritis	Spiegel et al., 2003	++	Uncertainty about the price of drug
	Colon Carcinoma	Earle et al., 2004	++	Uncertainty about the price of the drug
	Refractory epilepsy	Forbes et al., 2003	+	Uncertainty about the cost of device
	Progressive relapsing	Touchette et al., 2003	+	Uncertainty about the acquisition and administration cost of therapy
	Inguinal hernia	Stylopoulos et al., 2003	++	Uncertainty about the ambulatory facility cost
	Meningeal signs	Oostenbrink et al., 2002	+	Uncertainty about the treatment cost
	Hemophilia	Miners et al., 2001	++	Uncertainty about the clotting factor unit cost
	Heart burn	Heudebert et al., 2000	+	Uncertainty about the medication cost
	Breast cancer	Hayman et al., 2000	+	Uncertainty about the treatment cost
		Hutton et al., 1996	+	Uncertainty about the treatment cost
		Verma and Rocch, 2003	+	Uncertainty about the treatment cost
	Mild thyroid failure			
	Renal transplantation	Keown et al., 2001	++	Uncertainty about the length of hospitalization

Table 3.4 (Continued)

Factor	Disease	Reference	Magnitude of impact on cost–utility ratio*	Hypothesis or justification provided by author(s), if any
Treatment cost				
	Groin hernia repair	Medical Research Council Laparoscopic Groin Hernia Trial Group, 2001	++	Uncertainty about the cost of laparoscopic hernia repair, which might be due to theatre time and increased equipment and sterilization costs
	Parenteral nutrition program	Detsky et al., 1986	+	Uncertainty about the cost of alternative treatment
	Chronic myeloid leukemia	Gordois et al., 2003	+	Hypothesis about the drug price (10% to 50% discount)
	Lung cancer	Lievens et al., 2005	+	Uncertainty about the cost of Continuous Hyperfractionated Accelerated Radiotherapy (CHART)
Incorporation of indirect cost	Not applicable (review)	Meltzer D, 1997	N.A.	Uncertainty about the magnitude of future indirect cost
	Diabetes	Meltzer D et al., 2000	+	Uncertainty about the magnitude of indirect costs
	Gallstone disease	Cook et al., 1994	+	Uncertainty about the impact of indirect cost

*:
+: minor positive influence, indicating that increase in the factor would also result in minor increase in the cost-utility ratio but the impact was not large enough to change the conclusion regarding cost-effectiveness of the intervention ;
++: major positive influence, indicating that increase in the factor would also result in a substantial increase in the cost-utility ratio and the impact was large enough to change the conclusion regarding cost-effectiveness of the intervention;
−: minor negative influence, indicating that increase in the factor would result in minor decrease in cost-utility ratio but the impact was not large enough to change the conclusion regarding cost-effectiveness of the intervention ;
−−: major negative influence, indicating that increase in the factor would result in a substantial decrease in cost-utility ratio and the impact of the factor was large enough to change the conclusion regarding cost-effectiveness of the intervention;
N.A.: Not available (magnitude of the impact was not presented in the publication).
Unless indicated in the study, the threshold of cost-effectiveness was USD 50,000/QALY in this review.

3.3.3.5 Methodology-related factors

Cost-utility ratio was also affected by other methodology-related factors, such as discount rate (N=6), QALY elicitation method (N=2), statistical uncertainly (N=1), handling of confounding variables (N=1) and uncertainty about data source (N=1) (Table 3.5).

Discounting has been recommended as a standard practice in CUA to account for the time preference between now and future of both costs and benefits. According to the recommendations by the Panel of Cost-effectiveness in Health and Medicine, values in the future were suggested to be devalued by a constant annual percentage, equal for costs and effects (Siegel et al., 1996). Variations in discount rate have been shown to cause negative impacts on cost-utility ratios for both prevention programs and treatment interventions, the magnitude of which ranged from minor to major as shown in table 3.5 (Bonneux and Birnie, 2001; Seidenwurm et al., 2000; Brown et al., 2002). For example, in a hypothesized modeling study conducted by Bonneux and Birnie (2001), it was reported that an intervention eliminating cardiovascular disease cost 71,100 Euro/QALY with no discount. Applying different discount rates of 3% and 6%, the same intervention would cost 8,100 Euro/QALY (8.8 times less) and 1,100 Euro/QALY (65 times less), respectively. Moreover, in the meantime, the assumption of equally discounting cost and benefit has been much argued among various schools of thought and no consensus has been reached yet, which might also influence results of CUA (Mankiw, 2001; Viscusi, 1996; Cohen, 2003).

One of the most important components in CUA is the assessment of utilities.

QALY can be elicited using preference-based methods such as direct elicitation methods (time trade-off, standard gamble and rating scale) or by indirect methods using utility-based generic instruments (EQ-5D, SF-6D, etc) (Cook et al., 2001; Bravata et al., 2005). Although not directly proven by empirical studies, review articles focusing on QALY elicitation methods reported that different methods had resulted in different QALY values. It was commented that such difference would make results incomparable and may influence the magnitude and direction of results for evaluations (Richardson and Manca, 2004). In addition, Arneson and Trommald (2005) further proposed that other factors (e.g. age, gender, employment, etc) should also be taken into consideration even when comparing QALY values from the same method.

Uncertainties about Markov modeling have triggered heated discussions over recent years. In two sequential articles published by Nuijten et al. (2003; 2004), the authors generated the concern that cost-utility ratios produced by Markov modeling could be severely biased in the potential presence of statistical uncertainty (e.g. selection of p-value, type of estimation error), other potential confounding variables (e.g. covariance for time relationship between health state and the other explanatory variable, covariance between efficacy of the study drug and the other explanatory variable) or uncertainty about the data source (e.g. inclusion criteria, representativeness of the study population and source of the data whether from medical records, meta-analysis or others). However, such methodological thinking was strongly criticized by O'Hagan et al. (2005), that it might mislead and violate the current practices in handling uncertainties in modeling studies. In response to this,

Nuijten (2005) clarified that these concerns did exist, and more empirical studies should be conducted to verify such problems to further improve the current practices.

Table 3.5 Methodology-related factors that affect cost-utility ratio

Factor	Disease	Reference	Magnitude of impact on cost–utility ratio*	Hypothesis or justification provided by author(s), if any
Discount rate				
	Severe hemophilia	Miners et al., 2001	--	Different assumptions about the discount rate
	Visual loss	Brown et al., 2002	-	Different assumptions about the discount rate
	Cardiovascular disease	Bonneux and Birnie, 2001	--	Different assumptions about the discount rate
	Orbital foreign body screening	Seidenwurm et al., 2000	--	Different assumptions about the discount rate
	Orthoptic screening	König and Barry, 2004	-	Different assumptions about the discount rate
	HIV	Johnson-Masotti et al., 2000	-	Different assumptions about the discount rate
QALY elicitation method				
	Not applicable (review)	Arnesen and Trommald, 2005	N.A	Uncertainty caused by different values elicited from different QALY values
Data source				
	Not applicable (review)	Nuijten, 2004	N.A.	Uncertainty about the impact of data sources

Table 3.5 (Continued)

Factor	Disease	Reference	Magnitude of impact on cost–utility ratio*	Hypothesis or justification provided by author(s), if any
Handling of confounding variables				
	Not applicable (review)	Nuijten, 2004	N.A.	Uncertainty about the impact of confounding variables
Statistical uncertainty	Parkinson's disease	Nuijten, 2004	N.A.	Variation caused by of statistical uncertainty

*:
+: minor positive influence, indicating that increase in the factor would also result in minor increase in the cost-utility ratio but the impact was not large enough to change the conclusion regarding cost-effectiveness of the intervention ;
++: major positive influence, indicating that increase in the factor would also result in a substantial increase in the cost-utility ratio and the impact was large enough to change the conclusion regarding cost-effectiveness of the intervention;
−: minor negative influence, indicating that increase in the factor would result in minor decrease in cost-utility ratio but the impact was not large enough to change the conclusion regarding cost-effectiveness of the intervention ;
−−: major negative influence, indicating that increase in the factor would result in a substantial decrease in cost-utility ratio and the impact of the factor was large enough to change the conclusion regarding cost-effectiveness of the intervention;
N.A.: Not available (magnitude of the impact was not presented in the publication).
Unless indicated in the study, the threshold of cost-effectiveness was USD 50,000/QALY in this review.

3.4 Discussions

In this qualitative literature review, altogether 20 factors were identified that have shown to influence CUA results. They could be further grouped into five categories as treatment-related (duration, efficacy, and frequency of treatment), disease-related (severity level, risk level, incidence rate, prevalence rate, disease progression rate and survival length), patient-related (age, gender, race/ethnicity and compliance), cost-related (treatment cost, incorporation of indirect cost), and methodology-related factors (discount rate, QALY elicitation method, statistical uncertainty, handling of confounding variables and reliability of data source). Although most of the CUA studies included in the literature review were conducted in Western countries, yet it is highly possible that those 20 factors could potentially affect the CUA results in Asia.

To the best of our knowledge, this review article was the very first dedicated to identify all the influential factors of CUA in a qualitative and systematic way**.** The findings provide important implications and suggestions to further improve the robustness of CUA in the following perspectives:

First, the current list of 20 factors provides a comprehensive (yet definitely incomplete) checklist based on the current literature for both healthcare researchers and decisions-makers when conducting and interpreting CUA results. It is worth noting that the selection and investigation of influential factors should be adjusted according to the real scenario of the intervention and study population. In some cases, there might be no uncertainty about a few of the factors, while in other cases, there

could be other factors beyond the current list that might impose potential influence on CUA results. Hence, the current checklist is recommended to be used as a basic framework while allowing for further modifications.

Second, our results further emphasize the importance of conducting and reporting high quality sensitivity analysis; failure of which has actually been identified in a substantial proportion of CUA literatures (Schakman et al., 2004). As found in this review, cost-utility ratio may be subject to drastic changes in the presence of uncertainties induced by various parameters. Therefore, it is of great importance to identify all the sources of uncertainties in the first place. Then, as a good practice, extensive sensitivity analysis should be performed accordingly to well explore the impact of uncertainties to increase robustness of the results. If the uncertainty could not be investigated due to any restrictions, it should also be well recognized and documented to raise full awareness.

Last but not least, we would like to further point out that the factors in the current checklist might not be mutually independent. For example, efficacy of treatment was reported to be influenced by ethnicity, which could also influence other disease-related factors like incidence rate, risk level, prevalence rate etc (Kumar et al., 2006; G.ryskiewicz et al., 2006; Claudio et al., 2006; Kanuha, 2000). Therefore, due to such potential inter-correlations among various variables, the exploration of collective impact of these factors would be better off studied using multi-way sensitivity analysis to control for confounding effects in future studies.

In the meantime, some potential limitations of this literature review should also be recognized. First, although the current literature review was based on extensive search among several most comprehensive health-economics-related databases and top journals, it was possible that some other qualified articles published in other sources might not be retrieved. Second, the magnitude of minor or major impact as defined in the current review might not be accurate enough. For illustration purposes, we used a threshold of USD 50,000/QALY to differentiate whether the intervention was cost-effective for most of the cases, unless another specific threshold was mentioned in the publication. As such cut-off point is still in debate across countries or diseases, we suggested interpreting the magnitudes as presented in tables with caution.

3.5 Conclusion

The current list of factors could be used as a basic checklist when identifying sources of uncertainties and performing sensitivity analysis in CUA. Further empirical and theoretical studies were suggested to continuously provide better understanding of other potential influential factors so as to improve the robustness of CUA in medical decision-making.

Chapter Four

Is Diabetes Knowledge Associated with Health Utility Values among Subjects with Diabetes? A Preliminary Study among English-speaking Diabetic Subjects in Singapore

4.1 Introduction

In the disease management value chain, patients should not be merely recognized as recipients, who will get the estimated value of outcomes based on a certain amount of resource input. In fact, patients should be regarded as stakeholders, whose active participation in the disease management program would make a positive difference in the overall cost-effectiveness of the program and ensure the long-term optimization of the healthcare delivery process.

Such active participation has been commonly referred to as "patient empowerment", which emphasized the point that patient should assume their responsibility for their own disease and treatment; by incorporating a series of attitudinal and behavioral changes, these "empowered" patients would be more knowledgeable about, satisfied with and more committed to the treatment to achieve the best possible outcomes (Steele et al., 1987).

In the current chapter, we would focus on one of the most sought-after patient empowerment strategies, that is, disease knowledge and use diabetes population as an illustrative example.

Several published studies have shown that diabetes knowledge was positively associated with more frequently performed self-care behaviors (van den Arend et al., 2000) and better metabolic control (Beeney and Dunn et al., 1990; Nicolucci et al., 2000; Persell et al., 2004). It has also been reported that better metabolic control was associated with better Health-related Quality of Life (HRQoL) of diabetic patients (Vanelli et al., 2003; Mortensen et al., 2002; Hoey et al., 2001). However, relatively

little is known regarding the impact of diabetes knowledge on HRQoL. Furthermore, the impact of diabetes knowledge on health utility scores has not been explored either, the results from which would provide further rationale of whether disease knowledge should be incorporated as an influential factor when assessing the efficiency of disease management programs with cost-utility analysis.

Hence, in this preliminary study, our primary objective was to investigate the correlation between diabetes knowledge with health utility scores. In the mean time, we would also present the findings of the association between diabetes knowledge and HRQoL as a secondary objective.

4.2 Methods

4.2.1 Subjects and study design

The cross-sectional study was undertaken at the World Diabetes Day 2004 commemorative event organized by Diabetic Society of Singapore (DSS) at the exhibition hall of a major shopping center in Singapore on October 17, 2004. The aim of the event was to enhance people's knowledge of diabetes, preventive methods and other disease-related information as well as to provide free screening tests of cholesterol, blood pressure, body fat, HbA1c etc. Participants were approached by research assistants to identify whether they were eligible for the current study. Inclusion criteria were English-speaking diabetic patients aged 21 and above, who were able to complete questionnaires in English without any assistance. The purpose of the study was explained to potential subjects by the research assistants. All subjects who agreed to participate in the study were asked to sign a written informed consent.

Subjects were then asked to complete a booklet containing a background datasheet, diabetes knowledge test sheet, and pages of health utility and HRQoL measures. The background datasheet collected information on patient information including socioeconomic variables (age, gender, ethnicity, years of education, dwelling type, and working status) and related medical conditions (smoking history, presence of acute medical conditions in the past 4 weeks, presence of other chronic medical conditions and presence of diabetes complications). Patients' knowledge was tested by the General Diabetes Knowledge Test (GDKT). Their health utility scores were assessed by two generic preference-based utility measures, namely, EQ-5D and SF-6D. The disease-specific HRQoL measure, namely, the Audit of Diabetes-dependent Quality of life (ADDQoL) was used to assess patients' HRQoL.

4.2.2 Instruments

4.2.2.1 The General Diabetes Knowledge Test

General Diabetes Knowledge Test (Wee et al., 2005) is a 36-item questionnaire covering 6 content areas in diabetes prevention and control, namely, general knowledge of diabetes, risk factors of diabetes, symptoms of diabetes, complications of diabetes, treatment and management of diabetes and monitoring of diabetes (as shown in Appendix 4.1). It was a newly developed questionnaire by our research group with the aim to evaluate outcomes of diabetes education among general public including patients and non-patients, as such tool was lacking before. It has proven to be a valid and reliable measure. Internal consistency of GDKT was high (Kuder-Richardson Formula 20 or KRF20=0.9289). Item difficulty ranged from 0.59 to 0.97 and was significantly different between subjects with and without diabetes for 8 items

($p<0.05$). Test-retest reliability was moderate (Intra-correlation coefficient, or ICC=0.54). Construct validity was demonstrated using a known-group approach where subjects with diabetes were expected to have better knowledge of diabetes (i.e. higher GDKT scores) than subjects without diabetes. Here, any response of "yes" is regarded as a correct answer and marked as 1 point, "no" or "unsure" is regarded as wrong answer and marked as 0. A raw score is obtained based on the sum of correct responses. Then it is converted to a final score of 0-100 according to percentage. A higher score indicates better general knowledge of diabetes.

4.2.2.2 EQ-5D

Please refer to Section 2.2.3.2 for a detailed description of EQ-5D.

4.2.2.3 SF-6D

Please refer to Section 2.2.3.3 for a detailed description of SF-6D.

4.2.2.4 ADDQoL

Please refer to Section 2.2.3.1 for a detailed description of ADDQoL.

4.2.3 Statistical Analysis

Data were entered into an Excel spreadsheet (Microsoft Corporation, Redmond, Washington) and analyzed using the Statistical Package for the Social Sciences, version 12.0 (SPSS Inc, Chicago, Ill). Subjects with ≥10 missing values in GDKT or ≥6 missing items in ADDQoL or ≥ 1 missing value in EQ-5D were excluded from all analyses. Data with a normal distribution were reported as mean (standard deviation

or SD). Otherwise, the median (interquartile range or IQR) was reported. Descriptive analysis was used to characterize sociodemographics (age, gender, ethnicity, education level, smoking status, work status and dwelling type), medical information [type of diabetes, presence of acute disease(s) and co-morbidities and diabetes-related complications,] and scores of GDKT, EQ-5D, SF-6D and ADDQoL.

Relationships between various external variables and utility scores and HRQoL were explored by univariate analyses using Mann-Whitney or Kruskal-Wallis tests for categorical independent variables, or Spearman's correlation for continuous independent variables. Correlations between GDKT score and utility and HRQoL scores were studied in two steps, first with bivariate Spearman's correlation and then with partial correlation after adjustment for other potentially significant variables in the univariate analysis ($p<0.1$). Magnitude of correlation was interpreted according to the criteria proposed by Guyatt et al. (1999): less than 0.2 as very weak, over 0.2 but less than 0.35 as weak, over 0.35 but less than 0.5 as moderate, and over 0.5 as strong. If there was a statistically significant correlation ($p<0.05$), two-step multiple regression models were also constructed to further study the potential impact of disease knowledge on HRQoL or utility scores. In all the models, utility or HRQoL score was analyzed as the dependent variable. In the first step, disease knowledge was included as the only independent variable. In the second step, the impact of disease knowledge was studied after adjustment for potential confounding external variables identified in the univariate analysis ($p<0.1$). Statistical significance level was set at $p<0.05$.

4.3 Results

4.3.1 Characteristics of the subjects

Altogether 458 eligible subjects were approached at the event, however most of them declined to participate in the study. Thus 56 subjects participated in the study (response rate=12.2%), of which 14 were dropped due to incomplete responses as defined previously, leaving 42 subjects with complete responses for analysis. 90% of the eligible subjects had more than six years of education. Apart from diabetes, 64.3% had at least one other chronic medical condition, with 40.5% suffered from at least one diabetes complication.

In the current study, patients were reported with high level of diabetes knowledge based on the GDKT, with a median (IQR) of 94 (89, 100). Health utility values as generated by SF-6D and EQ-5D were at the higher end [Median (IQR) of SF-6D: 0.80 (0.70, 0.95); Median (IQR) of EQ-5D: 1.00 (0.80, 1.00)]. However, the HRQoL as measured by the disease-specific ADDQoL indicated that most of the patients were in poorer health status [Median (IQR): 49.57 (38.53, 64.67)]. [See Table 4.1]

Table 4.1: Characteristics of study subjects and scores of diabetes knowledge, health utility and HRQoL

Subject characteristics (N=42)	N (%) unless specified otherwise
Age (years) [Mean (SD)]	53.5(9.64)
Female	18(45.0)
Ethnicity	
Chinese	36(85.7)
Indian	5 (11.9)
Types of diabetes mellitus	
Type I	13(31.0)
Type II	29(69.0)
Dwelling type	
Public housing	33(78.6)
Private housing	9(21.4)
Years of education	
≤ 6	4(10.0)
7-12	18(45.0)
> 12	18(45.0)
Working	23(54.8)
Smoking	3(7.1)
Presence of acute medical conditions*	23(54.8)
Presence of other co-morbidities†	27(64.3)
Prevalence of diabetes complications‡	17(40.5)
Median (interquartile range) General Diabetes Knowledge Test scores§	94 (89,100)
Median (interquartile range) EQ-5D scores	1.00 (0.81, 1.00)
Median (interquartile range) SF-6D scores	0.80 (0.70, 0.95)
Median (interquartile range) ADDQoL scores‖	49.57(38.53, 64.67)

* Having had at least one of the acute disease(s) in the following five categories: a running nose/ sore throat/ cough, vomiting/diarrhea, a headache lasting more than one day, sleeping disorder, body injury in the past four weeks.

†: Having had at least one of the chronic disease(s) in the following nine categories, other than their primary diagnosis of diabetes: hypertension, heart disease, stroke, asthma or other lung disease, cancer, rheumatism or back pain or other bone or muscle illness, mental illness, other chronic diseases.

‡:Having had at least one of the diabetes-related complication (s) in the following seven categories: eye disease, foot problems, kidney disease, heart disease, stroke, erectile dysfunction, neuropathy.

§:A 36-item diabetes knowledge test, with a score standardized to 0-100; The higher the score, the

better the knowledge level.
‖: A 19-item diabetes-specific HRQoL measure, with a score standardized to 0-100; The higher the score, the better the HRQoL.

4.3.2 Correlation of diabetes knowledge with health utility values and HRQoL

Table 4.2 presents both the unadjusted and adjusted correlation coefficients between scores of GDKT and health utility and HRQoL. It was found that none of the unadjusted correlation between GDKT with EQ-5D, SF-6D, or ADDQoL was statistically significant. The unadjusted correlation coefficients were 0.02 (p=0.881) with EQ-5D, 0.05 (p=0.760) with SF-6D, and 0.20 (p=0.203) with ADDQoL.

In the univariate analysis, after adjusting for the parameter(s) that may potentially confound the correlation between diabetes knowledge and utility values and HRQoL ($p<0.1$), the respective correlation coefficients were 0.01 (p=0.936) with EQ-5D, 0.07 (p=0.715) with SF-6D, and 0.30 (p=0.081) with ADDQoL. Although the correlation coefficients between diabetes knowledge with SF-6D and ADDQoL were increased after the adjustment, neither of the relationship was statistically significant in the current study.

4.3.3 Exploration on the impact of diabetes knowledge on the ADDQoL score

Although in the partial correlation analysis, the correlation between diabetes knowledge and ADDQoL score was not statistically significant, yet a trend of correlation was observed (p=0.081). Such lack of significance level could be due to the potential impact of small sample size on the statistical level; therefore, we decided to carry on the exploration the impact of diabetes knowledge on the ADDQoL score for further verifications.

It was found that in the first regression model with GDKT score as the only independent variable, it explained up to 5.0 % of the variances in ADDQoL score (p=0.084). However, in the second step, after adjusting for the other confounding variables as found in the univariate analysis, it was found that GDKT no longer added any more to the explanative power of the other three external variables (adjusted R square=0.13, p=0.059), namely, education level, presence of chronic disease(s) and presence of diabetes-related complication(s).

Table 4.2 Correlation between diabetes knowledge and health utility values and HRQoL scores*

	Unadjusted Spearman's correlation coefficient	p value	Adjusted Spearman's correlation coefficient†	p value
EQ-5D	0.02	0.881	0.01	0.936
SF-6D	0.05	0.760	0.07	0.715
ADDQoL	0.20	0.203	0.30	0.081

*; Diabetes knowledge is measured by the Generic Diabetes Knowledge Test (GDKT, score range 0-100)

†: In the partial correlation analysis, variables that are adjusted based on the univariate analysis ($p<0.1$) are as follows:

EQ-5D: Education level (p=0.05)

SF-6D: Education level (p=0.09), presence of chronic disease(s) (p=0.02), Age (p=0.05)

ADDQoL: Education level (p=0.01), presence of chronic disease(s) (p=0.09), presence of diabetes-related complication(s) (p=0.09)

4.4 Discussion

In this preliminary study among English-speaking diabetic patients, we found that the correlation between diabetes knowledge (measured by GDKT) and health utility values (measured by generic health utility instruments as EQ-5D and SF-6D) was rather minimal, suggesting the impact of diabetes knowledge on the health utility values was lacking. Although not statistically significant, diabetes knowledge was found to be weakly correlated with disease-specific HRQoL score as measured by

ADDQoL. However, after adjusting for other variables, diabetes knowledge had no further impact on ADDQoL scores. Albeit preliminary, our findings could help to raise or address the following important issues regarding diabetes knowledge, health utility and HRQoL assessment:

First, if EQ-5D and SF-6D are taken as two generic health utility measures, then the lack of correlation between diabetes knowledge and EQ-5D and SF-6D scores suggest that diabetes knowledge might not be an influential factor on the utility values in cost-utility analysis when utilities are indirectly elicited from utility-weighted generic health status measures. Hence, it seemed rational and justifiable not to spend extra time and efforts to examine patients' knowledge during the data collection for cost-utility analysis.

However, if we take EQ-5D and SF-6D as the generic HRQol measure (Coons et al., 2000), then it is interesting and important to note from HRQoL scores that generic HRQoL measures might not be as sensitive and responsive as the disease-specific instrument (ADDQoL as in the study), which further reinforces the advantage and need of using disease-specific instrument to measure HRQoL of those patients. Furthermore, it suggested that utilities directly generated from disease-specific health states might be more accurate to evaluate the effectiveness of the disease management programs compared with generic health states.

Second, although the adjusted correlation between diabetes knowledge and ADDQoL was weak and insignificant, it suggested that the increase in diabetes

knowledge still have a role to play in improving HRQoL. Moreover, apart from the usual aim of increasing disease knowledge, patient education programs could also incorporate other elements of patient empowerment strategies such as communicating effectively with healthcare professionals, cultivating a positive attitude towards the disease management, etc (Aujoulat et al., 2007).

In the mean time, several limitations of this preliminary study also need to be pointed out. First, as this study was a cross-sectional study by convenience sampling at just one event on the same day, substantial number of patients could not be recruited and this might have led to the insignificant relationship as analyzed. Further studies utilizing larger sample size by purposive sampling would be required to confirm the findings. Second, in the current study, patients were reported to have a quite high level of disease knowledge, which could be due to the fact that GDKT was developed based on the findings of a public survey on general diabetes knowledge and was aimed to be used among the general population (Wee et al., 2002; Wee et al., 2005). Hence, its power to assess more specific diabetes knowledge might not be strong enough and a more comprehensive diabetes knowledge test like the one developed and validated by Michigan Diabetes Research and Training Center (Fitzgerald et al., 1998) is suggested to be used to validate the current findings. Third, future studies incorporating the measurement of clinical endpoints to measure diabetes control (e.g. HbA1c) would help bridge the theoretical link between diabetes knowledge and HRQoL/health utility and further verify our findings.

4.5 Conclusion

This preliminary study showed that diabetes knowledge did not impact generic health utility values or generic HRQoL scores. Yet there was a weak yet insignificant correlation between diabetes knowledge and disease-specific HRQoL score. Further studies utilizing larger sample size would be required to confirm the observation.

Chapter Five

Exploring the Impact of Health Literacy on Utility Assessment and Health-Related Quality of Life (HRQoL) among Patients with Rheumatic Diseases

5. 1 Introduction

In general, health literacy is defined by the American Medical Association (AMA), as "a constellation of skills, including the ability to perform basic reading and numerical tasks required to function in the healthcare environment." (AMA, 1999). It has been regarded as a better endpoint to measure patients' literacy capabilities in a medical setting as compared with general literacy (Villaire and Mayer, 2007). With the emerging emphasis on patient empowerment strategies in disease management, adequate health literacy has become an increasingly important factor in ensuring the effectiveness of health education programs and the quality of communication between patients and physicians (Safeer and Keenan, 2005; Speros, 2005). Inadequate health literacy has shown to be associated with lower treatment compliance rates, poorer biomedical and functional outcomes, and increased burden of illness, especially among patients with chronic diseases including diabetes, asthma, etc (Andrus and Roth, 2002; Schillinger et al., 2002; Wolf et al., 2005; Howard et al., 2005).

Since the emergence of health literacy studies in the medical literature in the early 1990s, the problem of inadequate health literacy has been reported among patients with rheumatic diseases including rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE). Patients with rheumatic diseases with limited health literacy were more likely to misinterpret prescription labels, have poorer disease knowledge and more hospital visits (Larson and Schumacher, 1992; Gordon et al., 2002; Buchbinder et al., 2000; Hearth-Holmes et al., 1997). However, as highlighted in a recent review by Rudd et al. (2007), most of the studies on health literacy in Rheumatology have focused on the assessment of readability of health education

materials, health literacy levels of patients with rheumatic diseases, or the suitability of health education materials for the intended study population. That is, little is known about the potential impact of health literacy on the health-related outcomes of patients with rheumatic diseases.

In spite of biological or physiological differences among different types of rheumatic diseases, patients with these diseases share some common outcomes, specifically disabilities in their physical, mental and social functioning (Wright, 1990; Brand et al., 1992; Wolfe, 1999; Allaire, 2001). As biomedical measures sometimes may not sensitively indicate the improvement in symptoms and health status, Health-related Quality of Life (HRQoL) has been increasingly incorporated as a complementary and essential outcome measure to assess changes in the physical, psychological, social and somatic functioning and well-being of these patients (Ward, 2004; Brunner and Giannini, 2003).

Given the perceived challenge of inadequate health literacy in the self-management skills of rheumatic diseases, it is important to explore whether health literacy influences HRQoL as an outcome among patients with rheumatic diseases (Ramos-Remus et al., 2000). Moreover, health utility scores generated from generic, preference-based HRQoL measures such as EQ-5D and SF-6D have been widely used in cost-utility analyses to determine the cost-effectiveness of pharmacological and surgical interventions in rheumatology for approval and subsidy decisions (Kobelt, 2006; Kavanaugh, 2007). Therefore, the exploration of the impact of health literacy on utility assessment would also be of great interest in determining if it is necessary to

incorporate health literacy as a potential factor influencing the results of cost-utility analysis of disease management programs among these patients. To address these issues and to help bridge the gap between health literacy and outcomes research in patients with rheumatic diseases, the current study aimed to explore the impact of health literacy on the utility assessment and HRQoL among patients with rheumatic diseases.

5.2 Methods

5.2.1 Subjects and study design

Consenting English-speaking patients with rheumatic diseases without cognitive problems who were over 18 years old and seen at a tertiary referral centre in Singapore were recruited by purposive sampling in this Institutional Review Board (IRB) approved study. Eligible patients were first asked to read out the 66 medical terms in the Rapid Estimate of Adult Literacy in Medicine (REALM) to the interviewer. Patients were next asked to self-complete two generic health utility measures (EQ-5D and SF-6D), a generic HRQoL measure (SF-36) and a pre-tested questionnaire to obtain information on patient's characteristics (including age, gender, ethnicity, education level, working status, dwelling type, presence of acute disease(s) and co-morbidities, primary diagnosis, disease activity and severity based on 10 cm horizontal visual analogue scales scored by the patient's rheumatologist). If the eligible patient reported that he or she was illiterate, he/she was be given a REALM score of 0, and HRQoL and subject characteristics were obtained through an interview – however, all subjects studied had some degree of literacy.

5.2.2 Measures

5.2.2.1 REALM

REALM is a 66-item word recognition test to assess the ability of an adult patient to read common medical words and lay terms relating to body parts and illnesses (See appendix 5.1). It was designed to assist medical professionals in estimating a patient's literacy level so that oral instructions and written education materials could be appropriately provided. REALM takes about two to three minutes to administer and score. According to the scoring scheme (range 0-66), results of REALM should be categorized and interpreted as follows: (1) 0-18: Will not be able to read most low literacy materials; will need repeated oral instructions, materials composed primarily of illustrations, or audio or video tapes; (2) 19-44: Will need low literacy materials; may not be able to read prescription labels; (3) 45-60: Will struggle with most patient education materials; and (4) 61-66: Will be able to read most patient education materials. (Davis et al., 1991) Correspondingly, in this study, patients were categorized into two groups as low health literacy (0-60) or adequate health literacy (61-66) to reflect their ability to read materials below or above ninth grade level.

5.2.2.2 SF-36

The SF-36 is a multi-purpose, short-form generic HRQoL measure with 36 questions on eight subscales, namely, physical function (PF), role-physical (RP), bodily pain (BP), general health (GH), vitality (VT), social function (SF), role-emotional (RE), and mental health (MH), with higher scores (range 0–100) reflecting better perceived health. The validity of SF-36 has been demonstrated in the English-speaking population in Singapore (Thumboo et al., 2001).

5.2.2.3 SF-6D

Please refer to Section 2.2.3.3 for a more detailed description of SF-6D

5.2.2.4 EQ-5D

Please refer to Section 2.2.3.2 for a more detailed description of EQ-5D.

5.2.3 Statistical analysis

Data were entered into a Microsoft Excel spreadsheet (Microsoft Corporation, Redmond, WA) and analyzed using SPSS 12.0 (SPSS Inc., Chicago, IL). All tests were two-tailed and conducted at a significance level of 0.05. Data with a normal distribution were reported as mean (standard deviation or SD). Otherwise, the median (interquartile range or IQR) was reported. Descriptive analysis was used to characterize sociodemographics (age, gender, ethnicity, education level, work status and dwelling type), medical information [presence of acute disease(s) and co-morbidities, primary diagnosis, activity and severity of the diagnosis] and scores of REALM, EQ-5D and SF-6D, and SF-36. The sociodemographic and other characteristics of patients with adequate and inadequate health literacy were compared using Student's t-test for continuous variables with a normal distribution, Mann-Whitney U non-parametric test for variables without a normal distribution, and the chi-square test for categorical variables.

Relationships between various external variables and utility scores and HRQoL were explored by univariate analyses using Mann-Whitney or Kruskal-Wallis tests for categorical independent variables, or Spearman's correlation for continuous

independent variables. Correlations between health literacy level and health utility and HRQoL scores were studied in two steps, first with bivariate Spearman's correlation and then with partial correlation after adjustment for other potentially significant variables in the univariate analysis ($p<0.1$). Magnitude of correlation was interpreted according to the criteria proposed by Guyatt et al. (1999): less than 0.2 as very weak, over 0.2 but less than 0.35 as weak, over 0.35 but less than 0.5 as moderate, and over 0.5 as strong. If there was a statistically significant correlation ($p<0.05$), two-step multiple regression models were also constructed to further study the potential impact of health literacy on utility or HRQoL scores. In all the models, utility or HRQoL score was analyzed as the dependent variable. In the first step, health literacy level was included as the only independent variable. In the second step, the impact of health literacy level was studied after adjustment for potential confounding external variables identified in the univariate analysis ($p<0.1$). Statistical significance level was set at $p<0.05$.

5.3 Results

5.3.1 Subject characteristics

Of 220 eligible subjects, 90.5% (n=199) participated. The majority of the patients were female Chinese with more than 10 years of education [70.5% female, 74.0% Chinese and 79.5% with at least 10 years education]. The commonest primary diagnoses among participants were Rheumatoid Arthritis (21.0%), Osteoarthritis (17.5%), Systemic Lupus Erythematosus (11.5%) or a Spondyloarthropathy (11.5%), with 47% of patients having at least one co-morbidities.

Based on REALM scores, 112 patients (56.3%) were categorized as having adequate health literacy and 87 patients as having low health literacy. Table 5.1 shows the comparison of subject characteristics between the two groups. As compared with low health literacy subjects, subjects with adequate health literacy were older, had more years of education and a higher chance of living in private housing.

5.3.2 Comparison of health utility and HRQoL scores in patients with rheumatic diseases by health literacy levels

No statistically significant difference in either EQ-5D or SF-6D utility scores was found between patients with adequate health literacy and those with inadequate health literacy. As for HRQoL, the physical functioning score (PF) of patients with adequate health literacy was significantly higher than those with inadequate health literacy [Median (IQR) of PF: 80 (60, 90) vs. 70 (45, 85) respectively, p=0.006]. Table 5.2 shows that apart from the difference in PF scores between these groups, there were no other significant differences between the two groups in the remaining SF-36 domains.

Table 5.1: Subject Characteristics

	Median (interquartile range), unless otherwise specified （N=199）		
	Adequate health literacy (N=112)	Low health literacy (N=87)	P value
Age (years) [Mean (SD)]	48.5 (14.7)	43.7 (14.0)	0.044
Ethnicity (N, %)			0.805
Chinese	82 (73.2)	65 (74.7)	
Malay	11(9.8)	10 (11.5)	
Indian	16 (14.3)	10 (11.5)	
Female (N, %)	76 (67.9)	64 (73.6)	0.383
Education (N, %)			0.010
≤6 years	4 (3.6)	8 (9.2)	
7-12 years	46 (41.1)	49 (56.3)	
>12 years	61 (54.5)	30 (34.5)	
Working (N, %)	66 (58.9)	54 (62.1)	0.654
Housing (N, %)			0.001
Private	51 (45.5)	20 (23.0)	
Public (i.e. Government subsidised)	61 (54.5)	67 (77.0)	
Presence of acute medical conditions * (N, %)	79 (70.5)	64 (73.6)	0.638
Presence of co-morbidities † (N, %)	54 (48.2)	38 (43.7)	0.525
Primary diagnosis (N, %)			0.462
Osteoarthritis ‡	23 (20.5)	11 (12.6)	
Rheumatoid Arthritis‡	20 (17.9)	22 (25.3)	
Systemic Lupus Erythematosus‡	12 (10.7)	11 (12.6)	
Spondyloarthropathy§	13 (11.6)	10 (11.5)	
Others ‖	40 (35.7)	30 (34.5)	
Activity of primary diagnosis ¶	0.50 (0.20, 0.65)	0.50 (0.20, 1.00)	0.050
REALM score††	64 (63, 65)	57 (51, 60)	0.000

* Having had at least one of the acute disease(s) in the following five categories: a running

nose/ sore throat/ cough, vomiting/diarrhea, a headache lasting more than one day, sleeping disorder, and body injury in the past four weeks.
†: Having had at least one of the chronic disease(s) in the following nine categories, other than their primary diagnosis of the rheumatic disease: diabetes, hypertension, heart disease, stroke, asthma or other lung disease, cancer, rheumatism or back pain or other bone or muscle illness, mental illness, other chronic diseases.
‡ : Based on the American College of Rheumatology (ACR) classification criteria
§: Based on the European Spondylarthropathy Study Group (ESSG) criteria
||: Other diagnosis ncluded hypermobility syndrome, osteoporosis, other forms or arthritis (e.g. gout, reactive arthritis), soft tissue rheumatism, fibromyalgia, other connective tissues diseases (e.g. myositis, Sjögren's syndrome) and systemic vasculitis,
¶: Measured on a 10 cm horizontal visual analogue scale, from "inactive" (scored as 0) to "very active" (scored as 10)
**: Measured on a 10 cm horizontal visual analogue scale, from "mild" (scored as 0) to "very severe" (scored as 10)
††: The 66-item Rapid Estimate of Adult Literacy in Medicine (REALM) is scored based on the correct number of answers, from 0 to 66.

Table 5.2: Comparison of health utility and HRQoL scores of patients with rheumatic diseases by health literacy levels

Instrument	Health utility or HRQoL scores of patients by health literacy levels Median (interquartile range) (N=199) *		
	Adequate health literacy (N=112)	Low health literacy (N=87)	P value
EQ-5D	0.80 (0.73, 1.00)	0.80 (0.73, 1.00)	0.645
SF-6D	0.80 (0.70, 0.88)	0.76 (0.66, 0.89)	0.160
SF-36†			
PF	80.00 (60.00,90.00)	70.00 (45.00,85.00)	0.006
RP	75.00 (0.00,100.00)	75.00 (0.00,100.00)	0.705
BP	62.00 (51.00,84.00)	62.00 (41.00,74.00)	0.323
GH	67.00 (47.00,77.00)	62.00 (45.00,72.00)	0.300
VT	65.00 (55.00,75.00)	65.00 (50.00,75.00)	0.252
SF	81.25 (60.50,100.00)	87.50 (62.50,100.00)	0.895
RE	100.00 (66.67, 100.00)	100.00 (66.67,100.00)	0.808
MH	80.00 (68.00,88.00)	76.00 (60.00,88.00)	0.497

*: Based on the scoring scheme of REALM, scores between 0 and 60 (less than 9th grade) are categorized under "low literacy" and scores between 61 and 66 (over 9th grade) are under "adequate literacy".
†: PF (physical function), RP (role physical), BP (bodily pain), GH (general health), VT (vitality), SF (social function), RE (role emotional), MH (mental health)

5.3.3 Correlation between health literacy levels and health utility and HRQoL scores

Utility scores as measured by SF-6D and EQ-5D scores were not significantly correlated with health literacy in either bivariate or partial correlations which adjusted for the respective significant external variables identified in the separate univariate analysis (Table 5.3).

In the exploration of correlation between health literacy level with HRQoL score as measured by SF-36, it was found that in univariate analysis, health literacy level was significantly correlated with PF; however this correlation was considered weak (Spearman's correlation coefficient=0.20, p=0.006). Other variables significantly associated with PF scores in univariate analyses were age, education level, working status, presence of co-morbidities and activity of the primary diagnosis. After adjustment for these five external variables in the partial correlation between the health literacy level and PF, the Spearman's correlation coefficient was 0.22 (p=0.002). This suggests that patients with rheumatic diseases with inadequate health literacy were more prone to worse physical functioning after adjusting for the influence of these other variables. However, as for the remaining SF-36 domains, no statistically significant relationship was found by either bivariate or partial correlation analysis.

Table 5.3 Correlation between health literacy level and HRQoL and utility scores of patients with rheumatic diseases*

	Unadjusted Spearman's correlation coefficient	P value	Adjusted Spearman's correlation coefficient †:	P value
EQ-5D	0.03	0.645	0.02	0.771
SF-6D	0.10	0.160	0.08	0.259
SF-36‡				
PF	0.20	0.006	0.22	0.002
RP	0.03	0.705	0.03	0.714
BP	0.07	0.323	0.04	0.554
GH	0.07	0.300	0.04	0.548
VT	0.08	0.252	-0.01	0.887
SF	0.10	0.895	-0.04	0.580
RE	0.02	0.808	0.02	0.832
MH	0.05	0.497	-0.01	0.900

*: Based on the Rapid Estimate of Adult Literacy in Medicine (score range: 0-66), health literacy levels are categorized into "low health literacy" (0-60) and "adequate health literacy" (61-66).

†: In the partial correlation analysis, variables that are significantly associated with SF-36 and utility scores are adjusted as follows:

EQ-5D: ethnicity, presence of co-morbidities, presence of acute disease, severity of the primary diagnosis, age, working status;

SF-6D: presence of acute disease, activity of the primary diagnosis, severity of the primary diagnosis, working status

PF: age, education level, working status, presence of co-morbidities and activity of the primary diagnosis are adjusted.

RP: presence of acute disease, activity of the primary diagnosis, severity of the primary diagnosis;

BP: presence of co-morbidities, activity of the primary diagnosis, severity of the primary diagnosis, working status;

GH: presence of co-morbidities, presence of acute disease

VT: age, presence of acute disease, activity of the primary diagnosis

SF: gender, presence of acute disease, activity of the primary diagnosis, severity of the primary diagnosis

RE: severity of the primary diagnosis, working status

MH: presence of acute disease, age, housing

‡ : PF (physical function), RP (role physical), BP (bodily pain), GH (general health);, VT (vitality), SF (social function), RE (role emotional), MH (mental health)

5.3.4 Impact of health literacy level on Physical function

As health literacy level was found to be only significantly associated with PF, further exploration of the impact of health literacy on PF was performed using multiple linear regression (MLR) models (Table 5.4). When health literacy level was the only independent variable (Step 1), it explained up to 3.6% of the variance in PF. When the five potential influential external variables (age, education level, working status, presence of co-morbidities and activity of the primary diagnosis) significantly associated with PF were adjusted for (Step 2), health literacy level was found to independently and significantly explain up to 3.7% of the variance in PF, with the other five external variables explaining up to 20.3% and the total of six variables explaining up to 24.0% of the variance in PF. Furthermore, based on the regression coefficient, it was also found that patients who were younger, working, having lower disease activity and with adequate health literacy were more likely to have better physical functioning.

Table 5.4 Impact of health literacy level on the physical functioning of patients with rheumatic diseases using multiple linear regression models

		Unstandardized regression coefficient (95% confidence interval)	Standardized Regression coefficient	p value
Step1 (unadjusted)	Adequate health literacy*	4.00 (1.27, 6.73)	0.20	0.004
Step2 (adjusted)	Adequate health literacy	4.14 (1.54, 6.74)	0.21	0.002
	Age (per year)	-0.18 (-0.28, -0.08)	-0.26	0.001
	Working	3.14 (0.47, 5.81)	0.16	0.022
	Presence of co-morbidities	0.43 (-2.28, 3.13)	-0.02	0.757
	Disease activity†	-3.31 (-4.87, -1.75)	-0.30	0.000
	Education			
	>12 years	1.45 (-3.80, 6.69)	0.07	0.587
	7-12 years	-0.57 (-5.96, 4.82)	-0.03	0.835
	≤6 years‡	0	0	

*: Based on the Rapid Estimate of Adult Literacy in Medicine (score range: 0-66), health literacy levels are categorized into "low health literacy" (0-60) and "adequate health literacy" (61-66).
†: Measured on a 10 cm horizontal visual analogue scale, from "inactive" (scored as 0) to "very active" (scored as 10).
‡: Baseline control group.

5.4 Discussion

In this cross-sectional study among patients with rheumatic diseases, we found that health literacy level did not impact health utility scores measured using either SF-6D or EQ-5D. Furthermore, it was also found that health literacy level did not influence HRQoL in general. Although there was a statistically significant correlation between health literacy level and physical functioning which persisted after adjusting for other variables, the strength of this correlation was weak and had an explanatory power of less than 4%. No significant impact of health literacy was found on the remaining 7 SF-36 domains. To the best of our knowledge, this is among the first studies exploring the impact of health literacy on health utility and HRQoL assessment in patients with rheumatic diseases. Our findings may provide useful information and important implications for the health literacy issues on patients with rheumatic diseases in several ways, detailed below.

First, it would be of special interest to the medical decision makers to know that health literacy level was not an influential factor of the final utility scores of patients with rheumatic diseases measured by EQ-5D and SF-6D. Utility scores elicited from patients with rheumatic diseases have been used to determine the effectiveness or the further cost-effectiveness of different interventions for the purposes of treatment selection and intervention reimbursement (Homik and Suarez-Almazor, 2004). Our results are reassuring in this regard as they suggest that health literacy does not influence utility scores in patients with rheumatic diseases. Thus there is unlikely to be a need to intentionally stratify recruitment of subjects on the basis of their health literacy.

Second, the weak yet significantly positive association between health literacy and physical functioning raises potential concerns over the problem of inadequate health literacy for various stakeholders in the management of rheumatic diseases. As preservation of the physical function is among the top priorities in the long-term care of patients with rheumatic diseases, our results suggest that physicians need to make extra efforts to effectively convey medical instructions to those with limited health literacy and ensure their compliance and proper execution on a regular basis (Simon, 2004; Sutcliffe et al., 1999). Similarly, health education materials for such patients need to be developed in such a way that they are well comprehended by patients with inadequate health literacy to optimize their positive impact on physical functioning in such patients. From a patients' perspective, those with inadequate health literacy could be encouraged take a more active role in communicating with their health care providers to maximize the possibility of improving their health status. It was also interesting to note that the correlation between health literacy and HRQoL outcomes detected in the current study was actually much higher than those found in other studies on patients with other chronic diseases like cancer (no association), depression (no association) and asthma ($r=-0.02$, $p<0.01$) (Hahn et al., 2007; Lincoln et al., 2006; Mankuso et al., 2006).

Last but not least, the high prevalence (44%) of patients with inadequate health literacy as identified by REALM in the current study further contributes to the increase in awareness of health literacy problems in patients with rheumatic diseases. According to the earlier cross-sectional studies on patients with SLE or RA, the percentage of inadequate health literacy ranged from 10% to 48% (Larson and

Schumaker, 1992; Buchbinder et al., 2000; Buchbinder et al., 2006). Therefore, the administration of a health literacy test could be recommended as a one-time screening test in routine healthcare settings to identify patients with rheumatic diseases with limited health literacy so as to deliver more tailored consultation and education.(e.g. further explaining medical terms in layman's language and designing special written materials for patients with low health literacy).

Finally, we recognize several limitations of this study. First, because the current study population included patients belonging to different categories of rheumatic diseases, no disease-specific preference-based health utility or HRQoL measure was used. Although the SF-36 appears to be the best available generic measure to evaluate health status of patients with rheumatic diseases, it might not be as sensitive as some other disease-specific HRQoL measures to capture small but meaningful differences in certain domains of HRQoL affected by a given disease. Besides, there is ongoing debate regarding the accuracy and sensitivity of using REALM, a word recognition test, to evaluate the comprehension capability, that is, the functional health literacy level of patients (Friedman and Hoffman-Goetz, 2006). These could possibly account for the weak (yet statistically significant) correlation between health literacy level and physical functioning. Future research could investigate the magnitude of such relationships using a disease-specific HRQoL measure. For example, in osteoarthritis, the Western Ontario MacMaster questionnaire (WOMAC) could be administered (Bellamy et al., 1988). Second, some characteristics of the study population may potentially mask the correlation between health literacy level with other domains of HRQoL. Because the study population in our study was mainly composed of those

with mild severity and disease activity, their mental health and other HRQoL functioning might be less affected than patients with more severe conditions. Therefore, in order to verify the lack of correlation between health literacy and other HRQoL functioning, it would be helpful to conduct the similar study on patients with more diversified levels of disease activity and severity. Apart from clinical endpoints such as disease activity and severity, incorporating the measurement of compliance level would also help bridge the theoretical link between health literacy and HRQoL/health utility and further verify our findings.

5.5 Conclusion

This study showed that health literacy did not impact generic health utility scores or HRQoL in general, but had a weak impact on the physical functioning of patients with rheumatic diseases. In order to confirm and expand the results of the current study, it is suggested that further studies be conducted on patients with specific rheumatic diseases using disease-specific utility and HRQoL instruments.

Chapter Six

Development and Validation of a generic functional health literacy test (GFHLT)

6.1 Introduction

Compared with the general definition of "health literacy" by the American Medical Association (AMA) as mentioned at the beginning of last chapter, "functional health literacy" is a more specific term referring to the ability of understanding and capability of acting on health information (Andrus and Roth, 2002). With the increasing emphasis on patient-oriented disease management and preventive medicine, patients as well as the general public need to achieve an adequate functional health literacy level to perform corresponding tasks in order to improve their clinical outcomes.

In health care settings, patients have to be functionally literate to be able to read and comprehend prescription labels, interpret appointment slips, follow instructions for diagnostic tests, complete medical documents, etc (Parker et al., 1995; Baker, 1995; Speros, 2005; Marrow et al., 2006). Inadequate functional health literacy has been shown to be related with poorer health status, lower compliance rates, prolonged hospitalization and increased health care costs (Guerra et al., 2005; Howard et al., 2005; Rutherford et al., 2006; Schillinger et al., 2006).

Similarly, the general public should also demonstrate relatively sufficient health literacy levels to comprehend various health education materials in order to continuously improve their health-related quality of life (HRQoL), strengthen their disease prevention capabilities and achieve the cost-effectiveness of those programs. Due to the poor correlation between functional health literacy and highest education level, it is not possible to reliably predict one’s functional health literacy level simply

from his or her education achievement (Weiss, 2001; Kleinpeter, 2003).

As for the measurement of functional health literacy, it has been debated that the widely used word recognition test, the 66-item Rapid Estimate of Adult Literacy in Medicine (REALM) might not be accurate and sensitive enough to measure functional health literacy levels (Friedman and Hoffman-Goetz, 2006). It is because the REALM only asks respondents to pronounce words rather than test comprehension of the words, which may sometimes lead to either under-estimation or over-estimation of the actual functional health literacy level.

Based on the literature search in PubMed (up till March 2007), there were altogether three validated functional health literacy tests for use in primary health care settings, namely, the Newest Vital Sign (NVS), the Test of Functional Health Literacy among Adults (TOFHLA) and the Test of Functional Health Literacy in Dentistry (TOFHLiD) (Parker et al., 1995; Weiss et al., 2005; Gong et al., 2007). A brief description of these three tests is provided as follows: The NVS is based on a nutrition label that is accompanied by 6 questions and requires 3 minutes for administration, which was recommended by the authors for use as a quick screening test for limited literacy in primary health care settings (Weiss et al., 2005). The TOFHLA is based on actual hospital materials used in the United States, which consists of a 50-item reading comprehension and 17-item numerical ability test, taking up to 22 minutes to administer (Parker et al., 1995). Using TOFHLA as a template, the TOFHLiD is to measure functional oral health literacy, with a 48-item reading comprehension and a 12-item numerical ability test related to fluoride use and access to care (Gong et al.,

2007).

However, none of these three tests was shown to be readily applicable for use among the general public for the following reasons. It has been highlighted that NVS was indistinguishable from a standard adult literacy test used in the National Adult literacy survey (NALS). That is, NVS was not specially developed to test the functional health literacy. Besides, NVS tended to assess quite a high proportion of mental mathematical skills, which may not be used that often by the general public in routine health education programs (King, 2005; Bennet, 2005). According to the discussions of TOFHLiD by its own developers, TOFHLiD was not recommended to be used among the general public or patients in primary health care settings other than those in dentistry (Gong et al., 2007).

Compared with NVS and TOFHLiD, TOFHLA was considered as a gold standard for measuring functional health literacy. Yet some of its drawbacks could not be ignored. Its incorporation of "Medicaid" (a unique health insurance system in USA) as one reading comprehension passage in both TOFHLA and S-TOFHLA (an abbreviated version of TOFHLA) made it lack of face validity when applied to other English-speaking populations (Baker et al., 1999). Besides, its relatively long completion time of 20 minutes prevented it from being administered as a screening test, especially in a busy clinic setting.

As none of the functional health literacy tests demonstrated appropriateness for use among general public of English-speaking populations around the world, a new

functional health literacy test should be developed and validated to fill this need. The primary objective of the current study was to develop and validate a generic functional health literacy test (GFHLT) for use among the English-speaking general public. A secondary objective of the current study was to further test the reliability and validity of GFHLT among English-speaking rheumatic patients in Singapore, as an example of its application and usefulness in a real-life health care setting.

The recruitment of English-speaking Singaporeans in the current study for the development and testing of the GFHLT has its unique advantages. Singapore has an education system based on the UK model which incorporates elements of the USA model, and English is the administrative language of government and the language of instruction at the local educational institutes, providing great potentials of adapting the current GFHLT to be used among other populations who are educated in English as well. Another advantage is that English health education programs developed in Singapore have been usually served as good examples within the Asian region, suggesting further potentials of administering GFHLT to the other English-speaking populations.

6.2 Methods

6.2.1 Study design and subjects

The whole study was divided into two phases: GFHLT development and GFHLT validation. The study was approved by the Institutional Review Board of the National University of Singapore and the Singapore General Hospital.

6.2.1.1 GFHLT development

As measured by the Flesch-Kincaid grade level index, a readability level of ninth grade was preset for the testing material of GFHLT, because the ninth grade had been shown to be the minimum functional health literacy level for patients to comprehend the majority of the health education materials (Davis et al., 1991; Eysenbach et al., 2002; Davis and Wolf, 2004; Aldridge, 2004; National Center for Education Statistics, 2007). An extensive review of health education materials published in Singapore by the same index also revealed that the readability levels ranged from sixth to twelfth grade level, with an average of around the ninth grade.

Due to the demonstrated simplicity and reliability to test reading comprehension in the education settings, the Maze procedure (a modified cloze test procedure) was adopted as the guideline to develop GFHLT (Rye, 1982). To increase the content validity, the 138-word passage was selected from a health education material called "Guidelines on Taking Medicines", which was published by the Singapore Health Promotion Board (Guidelines on Taking Medicines, 2007). According to the Maze procedure, every fifth word was taken out and replaced with a multiple choice item, which consisted of the correct word, a semantically incorrect distracter and a syntactically incorrect distracter. Hence, the GFHLT is a 21-item self-administered test to assess the functional health literacy level of respondents (See Appendix 6.1 for the GFHLT). Each of the 21 items was scored as either "0" for incorrect or "1" for correct, resulting in a range from "0" to "21" for the final score of GFHLT. As a rule of thumb in the interpretations of maze test suggested by Rye (1982), percentage of correctness was converted and assigned into categories of 0%-59% (frustrational

reading level), 60%-79% (instructional reading level) and 80%-100% (independent reading level).

When the preliminary version was ready, GFHLT was reviewed by two linguistic experts in English (researchers from the Center for English Language Communications with Ph D qualifications, National University of Singapore) and two pharmacists (from the National University of Singapore) to check the content validity. A pilot test of interviews with 55 respondents aged over 18 was then carried out by convenient sampling within the National University of Singapore to assess face validity, content validity and time of completion of GFHLT. Further suggestions on potential modifications of the GFHLT were also gathered from the respondents.

6.2.1.2 GFHLT validation

After any necessary amendments, the finalized questionnaires were distributed by research assistants to a convenience sample of 200 eligible Singaporeans at local neighborhoods and community centers, and a purposive sample of 200 eligible Singaporean patients with rheumatic diseases at a rheumatology outpatient clinic from a local tertiary hospital. Eligible respondents should be English-speaking subjects aged over 18 and with proper eyesight to read the questionnaire. Retests of the GFHLT were posted to contactable respondents at the second week and were required to be completed and posted back within one month after the first test.

6.2.2 Measures

During the validation stage, TOFHLA was not included, due to its drawbacks

such as being too time-consuming and lack of content validity among Singaporeans. Despite the lack of sufficient accuracy in measuring functional health literacy, the Rapid Estimate of Adult Literacy in Medicine (REALM) (Davis et al., 1991) was incorporated so as to study the construct validity of the GFHLT, as REALM shares the general construct with GFHLT as being a health literacy test. In addition, during the validation phase, sociodemographic and medical related information, such as age, gender, education level, working status, dwelling type, presence of chronic disease(s), was collected. At the same time, generic HRQoL of the participating subjects was measured by EQ-5D.

6.2.2.1 REALM

Please refer to Section 5.2.2.1 for a more detailed description of REALM.

6.2.2.2 EQ-5D

Please refer to Chapter 2, Section 2.2.3.2 for a more detailed description of EQ-5D..

6.2.3 Statistical analysis

Data were entered into a Microsoft Excel spreadsheet (Microsoft Corporation, Redmond, WA) and analyzed using SPSS 13.0 (SPSS Inc., Chicago, IL). All tests were two-tailed and conducted at a significance level of 0.01. Mean (standard deviation or SD) was reported for data with normal distribution, otherwise median (interquartile range or IQR) would be shown. Descriptive analysis was used to summarize sociodemographics (age, gender, education level, working status and housing type), the presence of chronic disease(s), EQ-5D and REALM scores of the

respondents.

Item difficulty was calculated based on the formula of:

D (item difficulty) = C (number of correct answers)/ N (number of respondents).

The item difficulty index has a range between 0.0 (most difficult) to 1.0 (least difficult). Item discrimination was analyzed by the corrected item-total correlation coefficient (range from 0.0 (least discriminative) to 1.0 (most discriminative) with a figure of being equal to or above 0.4 indicating adequate item discrimination (Cronbach, 1951). A statistical comparison by Student's t test was also performed to explore the potential differences in item difficulties between the samples of general public and the patients with rheumatic diseases. Average completion time of GFHLT was also calculated to evaluate any potential response burden.

Comparisons were made by Student's t test for continuous variables or Chi-square tests for categorical variables. EQ-5D, REALM and GFHLT scores were compared by Wilcoxon rank test. Reliability of GFHLT was assessed in terms of internal consistency by Cronbach's alpha, of which 0.7 was regarded as adequate reliability (Cronbach, 1951). After adjustment for significantly different external variables between the two populations, construct validity was studied by the partial correlation coefficients to evaluate the convergent validity between GFHLT and REALM scores, and the divergent validity between GFHLT score and education level. Test-retest reliability was assessed by intraclass correlation coefficient between the two measurements taken at baseline and at one-month post-test. Magnitude of correlation was interpreted according to the criteria proposed by Guyatt et al. (1999): less than 0.2 as very weak, over 0.2 but less than 0.35 as weak, over 0.35 but less than 0.5 as

moderate, and over 0.5 as strong.

6.3 Results

6.3.1 Scale development

GFHLT was judged as content valid by the two linguistic experts and two pharmacists. During the interview with a total of 55 respondents in the pilot test, face validity and content validity of GFHLT were recognized as well. Only one minor change was made based on suggestions from 10 respondents that the items should be numbered and bolded to prevent any missing answer (Appendix 6.1). The average completion time of GFHLT was 3 (2, 5) [Median (IQR)] minutes.

6.3.2 Scale validation

6.3.2.1 Subject characteristics

Altogether 223 respondents from the general public and 200 rheumatic patients participated in the validation study of GFHLT. Table 6.1 presents the subject characteristics. The mean (SD) age of the two groups were 43.8 (14.2) of the general public and 46.6 (14.7) of the rheumatic patients, respectively. The two groups were not significantly different in terms of education level, with about half of the respondents had more than 12 years of education.

However, compared with the general public, the sample of rheumatic patients had significantly higher proportion of female respondents (70.5% vs. 49.3%, $p<0.01$), but fewer working population (60% vs. 70.9%, $p=0.02$) and worse HRQoL as measured using EQ-5D [Median (IQR): 0.80 (0.73, 1.00) vs. 1.00 (0.81, 1.00), $p<0.01$].

Interestingly, it was found that the health literacy level of rheumatic patients were significantly higher than that of the general public [Median (IQR): 62 (58, 64) vs. 56 (50, 62) by REALM, p<0.01; and 19 (18, 20) vs. 19 (16, 20) by GFHLT, p<0.01].

Table 6. 1: Subject Characteristics

	Median (Interquartile range), unless otherwise specified		
	General public (N=223)	Rheumatic patients (N=200)	P value
Age (years) [Mean (SD)]	43.8 (14.2)	46.6 (14.7)	0.054
Female (N, %)	110 (49.3)	141 (70.5)	<0.01
Education (N, %)			0.581
≤6 years	25 (11.2)	12 (6.0)	
7-9 years	30 (13.5)	28 (14.0)	
10-12 years	66 (29.6)	68 (34.0)	
>12 years	102 (45.7)	91 (45.5)	
Working (N, %)	158 (70.9)	120 (60.0)	0.019
Housing (N, %)			0.349
Private	89 (39.9)	72 (36.0)	
Public	131 (58.7)	128 (64.0)	
Presence of chronic medical conditions* (N, %)	102 (45.7)	93 (46.5)	0.876
EQ-5D score	1.00 (0.81,1.00)	0.80 (0.73, 1.00)	<0.01
REALM score†	56 (50,62)	62 (58,64)	<0.01
GFHLT score ‡			
Baseline score	19 (16,20)	19 (18,20)	<0.01
Scoring categories (N, %)			
0%-59%	15 (6.1)	3 (1.5)	
60%-79%	47 (21.1)	19 (9.5)	
80%-100%	161 (72.8)	78 (89)	
1-month retest score	Not applicable	19 (18,20)	
Completion time of GFHLT	3 (2, 5)	3 (3, 4)	0.460

*: Having had at least one of the chronic disease(s) in the following 9 categories: diabetes, hypertension, heart disease, stroke, asthma or other lung disease, cancer, rheumatism or back pain or other bone or muscle illness, mental illness, other chronic diseases.
†: The 66-item Rapid Estimate of Adult Literacy in Medicine (REALM) is scored based on the correct number of answers, from 0 to 66. Each item is scored as either "0" for incorrect answer or "1" for correct answer.
‡: The 21-item Generic Functional Health Literacy Test (GFHLT) is scored based on the correct number of answers, from 0 to 21. Each of the items is scored as either "0" for incorrect answer or "1" for correct answer.

6.3.2.2 Item difficulty, item discrimination and response burden

Table 6.2 shows item difficulty and item discrimination of GFHLT. The item

difficulty ranged from 0.7 to 0.9 of the general public and 0.5 to 0.9 of the rheumatic patients. Item nine [Directions (to/for/and) use] and item ten [Any activity, food and (others/ other / which) medicines to be avoided] were shown to be the comparatively most difficult items in both populations. In terms of differences in the item difficulty between the two populations, the majority of the items were not significantly different, except for item nine [Directions (to/for/and) use], item 11 [(How/ Any / An) side effects] and item 13 [Be sure you (track/ follow / ignore) the instructions] with a small yet significant difference ranging from 0.1 to 0.2.

Based on the corrected Pearson item-total correlation coefficient, it was found that except for item one [Listen carefully to the (speech / instructions / then)], all the other items have reached adequate item discrimination power with coefficients of being equal to or above 0.4. The average completion time of the GFHLT in median was three minutes for both populations, suggesting its appropriateness of being used as a screening functional health literacy test.

Table 6.2 Item difficulty and item discrimination of the GFHLT[*]

No.	Item	Item difficulty[†]			Item discrimination[‡]	
		General public (N=223)	Rheumatic Patients (N=200)	P value	General Public (N=223)	Rheumatic Patients (N=200)
1	Instructions	0.9	0.9	0.77	0.3	0.2
2	Label	0.9	0.9	0.45	0.4	0.4
3	Ask	0.9	0.9	0.56	0.4	0.4
4	Understand	0.9	0.9	0.46	0.4	0.4
5	They	0.8	0.8	0.07	0.4	0.4
6	Should	0.8	0.8	0.12	0.4	0.4
7	And	0.8	0.9	0.04	0.4	0.4
8	Purpose	0.8	0.9	0.15	0.5	0.4
9	For	0.7	0.5	<0.01	0.5	0.4
10	Other	0.7	0.7	0.19	0.5	0.4
11	Any	0.8	0.9	<0.01	0.4	0.4
12	Out	0.8	0.9	0.14	0.5	0.4
13	Follow	0.8	0.9	<0.01	0.4	0.4
14	Exceed	0.9	0.9	0.12	0.4	0.4
15	Treatment	0.8	0.9	0.02	0.4	0.4
16	Prescribed	0.8	0.8	0.88	0.4	0.5
17	Complete	0.9	0.9	0.56	0.5	0.5
18	Medicines	0.9	0.9	0.05	0.4	0.5
19	Longer	0.9	0.9	0.41	0.4	0.5
20	Check	0.8	0.9	0.01	0.5	0.4
21	Taking	0.9	0.9	0.04	0.4	0.5

[*]: The 21-item Generic Functional Health Literacy Test (GFHLT) is scored based on the correct number of answers, from 0 to 21. Each item is scored as either "0" for incorrect answer or "1" for correct answer.
[†]: Item difficulty is calculated based on the formula: D = C/N, where D= item difficulty, C=number of correct answers and N=number of respondents (range: 0.0 (most difficult) to 1.0 (least difficult) ;
[‡]: Item discrimination is calculated as the corrected item-total correlation coefficient (range: 0.0 (least discriminative) to 1.0 (most discriminative) with a figure of being equal to or above 0.4 indicating adequate discrimination.

6.3.2.3 Reliability and construct validity

Cronbach's alpha of GFHLT was 0.72 of the general public and 0.68 of the rheumatic patients, suggesting adequate reliability as a generic functional health literacy test. Based on a retest sample of 112 rheumatic patients (representing a response rate of 56%), the test-retest reliability of GFHLT was shown to be high, with

an intraclass correlation coefficient of 0.95. As the response rate of the general public was too low (9%), the test-rest reliability was not calculated for such population.

After adjustment for significantly different external variables such as gender, working status and HRQoL score, the convergent validity of GFHLT was shown by the strong partial correlation between scores of GFHLT and REALM [0.59 ($p<0.01$) of the general public, 0.72 ($p<0.01$) of the rheumatic patients]. Similarly, after adjustment for the same set of external variables, the divergent validity was demonstrated by the weak correlation between GFHLT score and education level [0.33 ($p<0.01$) of the general public, 0.28 ($p<0.01$) of the rheumatic patients].

6.4 Discussions

To the best of our knowledge, the 21-item GFHLT is the first test specially developed to screen the functional health literacy level of the general public. The face validity and content validity of GFHLT were well recognized by linguistic experts and pharmacists during the expert judgment and by another 55 eligible respondents from the general public during interviews. At the validation stage, GFHLT has further demonstrated its adequate reliability, good construct validities and sufficient item discrimination power in both study populations, the general public (N=223) and rheumatic patients (N=200) in Singapore. Furthermore, as GFHLT could be completed in about three minutes, it has great potential to be used as a screening test of functional health literacy in health education and clinical settings. Besides, GFHLT could also be incorporated into other disease management evaluation studies to assess the functional health literacy either as a dependent or as an independent variable with

minimal response burdens.

Apart from the key results, some of the other findings are also worth mentioning for a variety of interests. First, the results from our study support the observation that highest education level should not be used as a proxy for functional health literacy level due to their weak correlations. Second, we found that item difficulty of the majority of items was skewed towards easier items, with indices of over 0.5. Although such result could be explained by the similarly high-end skewed literacy levels of the respondents measured by both REALM and GFHLT, further refinement of the distracters in choices could be made to increase the difficulty levels to the recommended range of 0.3 to 0.7 or to an optimal of 0.5, if possible. The further improvement in the item difficulty would also be expected to increase the reliability level of the GFHLT (Aiken, 1997). Third, as the readability level of GFHLT was preset at the ninth grade, we suggested the ninth grade to be used as a reference point when categorizing respondent's functional health literacy capabilities into frustrational, instructional and independent levels.

We recognize several limitations of this study, which lead to suggestions for future studies and are thus presented together. Due to the poor response rate of the general public at retest, the test-retest reliability was not known on such population. Hence, this would need to be ascertained with further studies, and we suggest that tokens for respondents could be given out to increase response rate of the retest to prove the test-retest reliability of the GFHLT among the general public. Besides, due to practical constraints which did not allow implementation of probability sampling strategies, the

current sample included only a handful of respondents with very low literacy levels. It could be possibly due to the shyness and reluctance of them to participate in surveys as found in other studies (Dowse and Ehlers, 2005; Miller et al., 2007). Therefore, we would suggest more efforts to be made for recruitment of such respondents by purposive sampling in both the general public and rheumatic patients to explore the sensitivity of GFHLT in identifying respondents with frustrational functional literacy levels.

Despite the limitations, it is expected that the availability of GFHLT would help health care educators and practitioners develop more proper health education programs at compatible health literacy levels so as to enhance the overall cost-effectiveness of those disease management programs. Potential studies may include but are not limited to the exploration of impact of functional health literacy on a variety of outcomes such as compliance, health-related quality of life, cost-effectiveness of the programs, etc. Consequently, better health education outcomes would contribute to the increasing health care awareness, improved HRQoL and more empowered disease management capabilities of the populations (Hoffmann and Worrall, 2004;.Kleinbeck, 2005).

6.5 Conclusion

In conclusion, our results demonstrated that the 21-item GFHLT is a reliable and valid screening test to measure functional health literacy levels of the general public as well as patients in a health care setting. The availability of GFHLT would contribute to the more accurate assessment of the functional health literacy and further

exploration of its impact on the cost-effectiveness of disease management programs.

Chapter Seven

An Exploratory Study of Response Shift in Health-Related Quality of Life (HRQoL) and Utility Assessment among Patients with Osteoarthritis Undergoing Total Knee Replacement Surgery

7.1 Introduction

Originating from self-report evaluation studies on organizational and educational psychology, a phenomenon termed "response shift" has been increasingly reported in patient-reported outcomes (PROs) assessment among chronically or terminally ill patients over the last decade (Golembiewski et al., 1976; Zwiebel, 1987; Postulart and Adang, 2000).

In the healthcare arena, response shift has been regarded as an instinctive psychological mechanism of patients to adapt to the changes caused by illness or treatment (Wilson, 1999). Sprangers and Schwartz (1999) have defined response shift as a change in the meaning of one's self-evaluation of a target construct as a result of: (a) a change in the respondent's internal standards of measurement (scale recalibration); (b) a change in the respondent's values (scale reprioritization); or (c) a redefinition of the target construct (reconceptualization)." Although distinguished as three types of response shift, reconceptualization, scale recalibration and reprioritization are thought to occur in combination (Ahmed et al., 2005).

Health-related quality of life (HRQoL) measurement may be affected by response shift because it quantifies patient perceptions, which may change with time due to response shift. Paradoxes such as overestimation of health status or underestimation of treatment effects measured by HRQoL outcomes have been found across various patient groups including cancer (Breetvelt and Van Dam, 1991), stroke (Ahmed et al., 2005), mental illness (Schweickhardt et al., 2005), etc. A theoretical model has been built to illustrate the relationship between response shift and HRQoL,

in which “changes in an individual’s health status may prompt behavioral, cognitive and affective processes necessary for accommodating illness, which may be influenced by antecedents (e.g. sociodemographics, personalities, expectations, etc) of the individual; these processes have the potential to change an individual’s standards, values and conceptualization of HRQoL.”(Breetvelt and Van Dam, 1991). The presence of response shift calls into question the assumption that patients would perceive and value a self-reported item with entirely the same internal standards during longitudinal research. In other words, there may be situations where true change measured by HRQoL instruments may not be simply calculated as the difference between respective pre- and post-intervention test scores (Howard et al., 1979; Ahmed et al., 2004).

In various evaluations of disease management programs, pre- and post-intervention comparisons of HRQoL have been used as a standard method to evaluate patients’ improvement in both generic and disease-specific health status, and consequently to determine cost-effectiveness of the interventions (Bernhard et al., 2001; Fetterrolf et al., 2004). Due to the potential impact of response shift on such pre-post comparison outcomes, it is thus necessary to explore the presence of response shift and the magnitude of it in either HRQoL or utility assessment. In the current thesis, our study would use total knee replacement (TKR), a surgical intervention for managing the end-stage knee osteoporosis, as an illustrative example.

The exploration of response shift has become an emerging area in HRQoL research of surgical interventions (Thompson-Fawcett et al., 2000; Oort et al., 2005;

Roos et al., 2004). There is however limited information currently available on the impact of response shift in subjects undergoing TKR, with only one recent publication showing that response shift significantly affected postoperative function six months after TKR when measured using a disease-specific HRQoL questionnaire, the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) (Razmjou et al., 2006). However, the impact of response shift on generic HRQoL instruments, including preference based HRQoL instruments (e.g. the EQ-5D or SF-6D) is not known. If present, this may lead to inaccurate or even invalid results when these instruments are used in utility assessment in this situation. Additionally, evidence of its influence on TKR patients over a follow-up period longer than six months is also lacking. Neither is it clear whether response shift also affects comparisons between two postoperative time points for recovery assessment.

To address these gaps in the literature, the primary objective of the current study was to explore and compare the impact of response shift on HRQoL and utility scores measured by generic HRQoL instruments at baseline and six months after TKR when assessed 18-months after TKR. It was hypothesized that response shift at baseline would be larger than that at six months post TKR, given that there was no major intervention between six months and 18 months postoperatively. If response shift were demonstrated, potential demographic and health-related factors associated with response shift at that time point would be investigated. In addition, the agreement between SF-6D and EQ-5D in detecting response shift would also be explored. Based on a comparison study of EQ-5D and SF-6D across seven patient groups including osteoarthritis, it was hypothesized that correlation of response shift between the two

measures would be moderate as categorized by Cohen's criteria (a correlation coefficient within the range of 0.3-0.5) (Brazier et al., 2004; Cohen, 1988).

7.2 Methods

7.2.1 Subjects and study design

By purposive sampling, contactable consenting TKR patients without cognitive problems who underwent the TKR at the orthopedic surgery department a tertiary referral centre in Singapore in 2005 (the most established orthopedic surgery department in the country) were recruited in this Institutional Review Board (IRB) approved study. Due to difficulties in communication during the telephone survey (the third phase as mentioned below), dialect-speaking patients who could not speak either English or Mandarin Chinese (N=19) were excluded. In addition, patients undergoing any additional surgery during the study period would also be excluded to obviate any confounding physical and psychological impact caused by this additional surgery.

This IRB approved prospective study was carried out in three phases. Data for the first two phases were retrieved from an earlier IRB approved study, in which generic HRQoL and utility scores were determined by an interviewer using the SF-6D and EQ-5D at baseline before the surgery (pre-test 1), and using the SF-6D six months after surgery (pre-test 2) (Xie et al., 2007). Response shift was studied using the "then-test" approach in the third phase, in which eligible Mandarin- or English-speaking patients were interviewed through the telephone 18 months after their surgery. In this telephone interview, patients were asked to give their HRQoL scores for their current health status using both the SF-6D and EQ-5D (i.e. post-test scores).

They were also asked to give their HRQoL scores at baseline (i.e. then-test 1 scores) and six months after TKR (i.e. then-test 2 scores). The rationale of the then-test approach is that at post-test using the same measure, respondents will provide their retrospective judgment of the health status at baseline and six months using the same internal standard (Sprangers and Schwartz, 1999). In the scoring scheme of the then-test approach, response shift is calculated as the difference between pre-test and then-test scores for each time point assessed, in this case at baseline and six months after TKR. True change or adjusted treatment effect is calculated as the difference between respective post-test and then-test scores. The difference between respective post-test and pre-test scores was considered the observed change or unadjusted treatment effect (Brossart et al., 2002).

Additional data collected during the telephone survey included demographics (age, gender, education level, work status, dwelling type), medical information (presence of acute or chronic illnesses, past knee surgery, number of knees operated) and general satisfaction with knee surgery (on a 0-10 Likert scale, with 0 means "not satisfied at all" and 10 means "totally satisfied").

7.2.2 HRQoL Measures

7.2.2.1 SF-6D

Please refer to section 2.2.3.3 for a more detailed description of SF-6D.

7.2.2.2 EQ-5D

Please refer to Section 2.2.3.2 for a more detailed description of EQ-5D.

7.2.3 Statistical analysis

Data were entered into a Microsoft Excel spreadsheet (Microsoft Corporation, Redmond, WA) and analyzed using SPSS 13.0 (SPSS Inc., Chicago, IL). All tests were two-tailed and conducted at a significance level of 0.01 to reduce the possibility of spurious tests of significance due to multiple comparisons. Descriptive analyses were used to characterize demographics (age, gender, education level, work status, dwelling type), medical information (presence of acute or chronic illness, past knee surgery, number of knees operated) and general satisfaction with knee surgery (on a 0-10 Likert scale). Data with a normal distribution were reported as mean (standard deviation or SD). Otherwise, medians (interquantile range or IQR) were reported. Mann-Whitney tests were used to investigate whether there were any statistically significant differences between the respondent group and non-respondent group in terms of demographics, medical information and health status at baseline and six months after the surgery. Wilcoxon Signed Rank tests were used to assess the significance of differences between pre- and then-test scores, between true change and observed change at baseline and six months after surgery, between magnitude of response shift and between changes in levels of SF-6D domains of response shift at baseline and six months after surgery.

Relationships between response shift and external variables were investigated by univariate analyses using Mann-Whitney or Kruskal-Wallis tests for categorical independent variables, or Spearman's correlation for continuous independent variables. Independent variables with $p < 0.10$ in univariable analyses were then entered into respective multiple linear regression (MLR) models to explore factors

potentially impacting response shift. The dependent variable for each MLR was response shift at baseline or at six month after TKR, respectively. Due to the small number of subjects, we considered the results of MLR analysis as exploratory. The agreement between SF-6D and EQ-5D in detecting response shift was explored by Spearman's correlation and Bland-Altman plots.

7.3 Results

7.3.1 Response rate and subject characteristics

It was noted that none of the 117 identified patients underwent additional surgery during the study period. The response rate of in this 18-month follow up study was 63% (74 of 117 patients). Altogether 43 patients were not recruited for the following reasons: dialect-speaking (N=19), uncontactable (N=10), declined participation (N=5), cognitive problems (N=4), overseas residence (N=2), admission into hospital (N=1), deafness (N=1) and death (N=1). There was no statistically significant differences between responders and non-responders in terms of demographics (age, gender, education level, work status, dwelling type), medical information (presence of acute or chronic illness, past knee surgery, number of knees operated), and health status at baseline and six months after the surgery.

The majority of respondents were elderly women with few years of education and with chronic disease(s) [median (IQR) age: 68 (63, 76) years, 81% female, 92% with less than 12 years of education, 68% with at least one chronic disease]. Patients' satisfaction with the TKR was high, with a median (IQR) score of 8 (8, 9)]. More detailed subject characteristics are presented in Table 7.1.

Table 7.1: Subject Characteristics

	Median (interquantile range), unless otherwise specified (N=74)
Age (years)	68 (63,76)
Ethnicity (N, %)	
Chinese	66 (89.2)
Malay	3 (4.1)
Indian	4 (5.4)
Female (N, %)	60 (81.1)
Education (N, %)	
≤6 years	49 (66.2)
7-12 years	19 (25.7)
>12 years	6 (8.1)
Working (N, %)	10 (13.5)
Housing (N, %)	
Private	15 (20.3)
Public	59 (79.7)
Presence of acute disease(s) * (N, %)	53 (71.6)
Presence of chronic disease(s)† (N, %)	50 (67.6)
Past knee surgery (N, %)	7 (9.7)
Number of knee(s) operated on at baseline(N, %)	
1 knee	59 (79.7)
2 knees	15 (20.3)
Satisfaction with the operation(s) ‡	8 (8,9)
SF-6D index for current health status at 18 months	0.77 (0.66, 0.90)
EQ-5D index for current health status at 18 months	0.87 (0.71,1.00)

*: Having had at least one of the acute disease(s) in the following 5 categories: a running nose/ sore throat/ cough, vomiting/diarrhea, a headache lasting more than 1 day, sleeping disorder, body injury in the past four weeks.

†: Having had at least one of the chronic disease(s) in the following 9 categories: diabetes, hypertension, heart disease, stroke, asthma or other lung disease, cancer, rheumatism or back pain or other bone or muscle illness, mental illness, other chronic diseases.

‡: Self-reported satisfaction with the total knee replacement was measured on a 11-point Likert Scale from 0 (totally unsatisfied) to 10 (totally satisfied).

7.3.2 Presence and impact of response shift

As seen in Table 7.2, median (IQR) SF-6D scores of then-tests at baseline [0.48 (0.42, 0.49)] and six months after TKR [0.72 (0.66, 0.79)] were significantly

different from the respective pre-test scores [0.61 (0.58, 0.68) at baseline, $p<0.001$; 0.69 (0.63, 0.72) at six months after TKR, $p<0.001$], indicating the presence of response shift at both time points. Interestingly, response shift at baseline [0.14 (0.08, 0.20)] was not only significantly larger than that at six months after TKR [-0.05 (-0.14, 0.00), $p<0.001$], but also in opposite directions. When measured by EQ-5D, significant difference was also detected between pre-test and then-test scores at baseline [0.69 (0.17, 0.73) for pre-test vs. -0.18 (-0.23, 0.00) for then-test, $p<0.001$].

Thus when response shift was considered in studying the impact of TKR, the adjusted improvement in health status became significantly greater between the pre- and six-month postoperative period [true change of 0.30 (0.18, 0.39) by SF-6D and 0.72 (0.22, 0.91) by EQ-5D, $p<0.001$]. On the contrary, after adjustment, the treatment effect between the periods of six months and 18 months after surgery became quite minimal, though statistically significant [true change of 0.03 (0.00, 0.09) by SF-6D, $p<0.001$]. (Table 7.2)

Table 7.2. Horizontal comparisons between pre- and then-test, and between true change and observed change at baseline and six months after total knee replacement *

Baseline									
	N	Pre-test 1†	Then-test1 ‡	Magnitude of Response Shift§	p value	Post-test at 18 months	Observed change ‖	True change ¶	p value
Baseline (SF-6D)	71	0.61 (0.58,0.68)	0.48 (0.42,0.49)	0.14 (0.08, 0.20)	<0.001	0.77 (0.66,0.90)	0.16 (0.02,0.26)	0.30 (0.18,0.39)	<0.001
Baseline (EQ-5D)	68	0.69 (0.17,0.73)	-0.18 (-0.23,0.00)	0.72 (0.22, 0.91)	<0.001	0.87 (0.71,1.00)	0.27 (0.02,0.62)	0.95 (0.65,1.16)	<0.001
At 6 Months									
	N	Pre-test2 **	Then-test2 ††	Magnitude of Response Shift	p value	Post-test at 18 months	Observed change	True change	p value
Six months after the surgery (SF-6D)	62	0.69 (0.63,0.72)	0.72 (0.66,0.79)	-0.05 (-0.14,0.00)	<0.001	0.77 (0.66,0.9)	0.10 (0.03,0.21)	0.03 (0.00,0.09)	<0.001

*: Data are reported as Median (interquantile range).

†:Pre-test 1: measure taken at baseline, just prior to the TKR surgery.

‡: Then-test1: retrospective measure of baseline score taken at 18 months after the TKR surgery

§: Magnitude of response shift= respective (pre-test score- then-test score).

Please note that the result was based on the statistical analysis of response shift of 74 patients, so it might be slightly different from the result of a simple subtraction.

‖: Observed change= respective (post-test score - pre-test score).

¶: True change=respective (post-test score - Then-test score).

**:Pre-test 2: measure taken at six months after the TKR surgery.

††: Then-test 2: retrospective measure of score of six-month postoperative score taken at 18 months after the TKR surgery.

Similarly, when the magnitude of response shift was further studied using individual SF-6D items [Table 7.3], it was found that the degree of impairment in all six domains (each measured by one item) as were rated as more severe during the then-test at baseline. A similar situation was also found in the then-test six months after TKR, except for the domain of "vitality" where the direction of response shift was towards less severe reduction, in contrast with that observed for the other domains.

Interestingly, contrary to the quantitative data showing the presence of response shift provided by patients, these same subjects generally did think that their then-test and pre-test ratings were similar. At the end of the telephone survey, when asked whether their then-test ratings were different from pre-test ratings, 70 out of 74 patients thought that these would be similar for both time points assessed (i.e. at baseline and six month postoperatively). At baseline, the magnitude of response shift of patients who thought their scores were similar (N=70) [by EQ-5D: 0.72 (0.25, 0.90); by SF-6D: 0.15 (0.08, 0.20)] was slightly larger than patients who were actually aware of the difference (N=4) [by EQ-5D: 0.57 (0.04, 1.02); by SF-6D: 0.12 (0.07, 0.25)]. However, there seemed to be no obvious difference between the two groups regarding response shift at six months after TKR [SF-6D scores of -0.05 (-0.14, -0.00) vs. -0.05 (-0.19, -0.05) respectively].

Table 7.3 Response shift in domains of SF-6D at baseline and six months after total knee replacement *

SF-6D domains†	Median (Interquantile)		P value
	0-month	6-month	
Physical functioning (6 levels)	-1.00 (-1.00,0.00)	-1.00 (-1.00,0.00)	0.059
Role limitations (4 levels)	-2.00 (-2.00,0.00)	-0.50 (-1.00, 0.00)	<0.001
Social functioning (5 levels)	-2.00 (-3.00,0.00)	-1.00 (-1.00, 0.00)	0.002
Pain (6 levels)	-1.00 (-2.00, 0.00)	0.00 (-1.00, 0.00)	<0.001
Mental health (5 levels)	-2.00 (-2.00, -1.00)	0.00 (-1.00, 0.00)	<0.001
Vitality (5 levels)	-1.00 (-2.00, 0.00)	1.00 (0.00, 2.00)	<0.001

*: Response shift= Pre-test level - Then-test level.
†: Each domain is measured by numeric levels from "1" onwards, with higher number indicating worse condition.

7.3.3 Influence of external variables on response shift

Univariate analysis indicated that response shift at baseline measured using the SF-6D was significantly influenced by education level (p=0.042), working status (p=0.050) and presence of chronic disease(s) (p=0.037). Interestingly, patients with less education and chronic disease(s) who were not working experienced a larger degree of response shift, compared with those having higher education level and a job but no chronic disease. However, the six-month postoperative response shift measured using the SF-6D was only significantly influenced by education level (p=0.004). In contrast with the findings at baseline, patients with more education experienced a larger degree of response shift. No further variables were incorporated into either regression model, as their respective p values in the univariate analysis were all more than 0.1.

The exploratory MLR analysis (Table 7.4) suggested that the combination of education level, working status and presence of chronic disease accounted for 8% of the variance in baseline response shift (p=0.05). At six-months postoperatively, educational level was the only external variable incorporated, accounting for 16% of the variance (p<0.01).

Table 7.4 Influence of external variables on respective response shift at baseline and six months after total knee replacement

MLR on baseline response shift *			
Independent variable(s) †	Regression coefficient (95% confidence interval)	P value	Adjusted R square
Education level ‡			
7-12 years	-0.04 (-0.10, 0.02)	0.179	
>12 years	-0.08 (-0.17, 0.00)	0.068	0.08
Working status ‡	-0.02 (-0.10, 0.06)	0.608	
Presence of chronic disease(s) ‡	0.41 (-0.01, 0.10)	0.110	
MLR on response shift at six months after total knee replacement			
Independent variable †	Regression coefficient (95% confidence interval)	P value	Adjusted R square
Education level ‡			
7-12 years	-0.05 (-0.10, 0.00)	0.055	0.16
>12 years	-0.12 (-0.20, -0.05)	0.001	

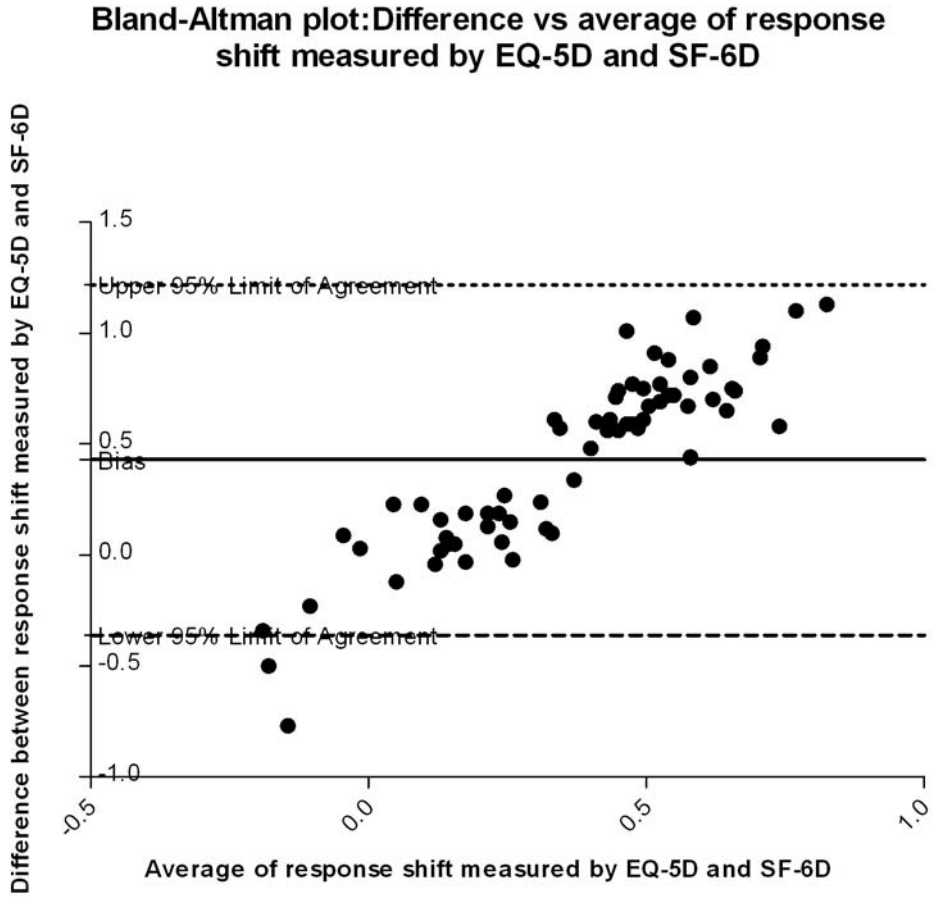
*: Multiple Linear Regression Model;
Response shift in the table was measured by SF-6D.
†: Only external variable(s) with p < 0.05 in univariable analyses were incorporated into (MLR) models as independent variable (s). No further external variables had p value <0.1.
‡: Reference categories of education level, working status and presence of chronic disease(s) were ≤6 years of education, not working, and with no chronic disease(s), respectively.

7.3.4 Systematic difference between SF-6D and EQ-5D of detecting response shift

In the current study, the EQ-5D was found to have a significantly larger magnitude of response shift at baseline [0.72 (0.22, 0.91)], as compared to the SF-6D [0.14 (0.08, 0.20)]. A moderate correlation [Spearman's correlation coefficient of 0.43 (p<0.001)] in response shift at baseline for these two measures was found as hypothesized. Further comparison of the degree of response shift using a Bland-

Altman plot (Figure 7.1) also showed moderate agreement between the two measures, with most of the data points falling in between the lower and upper 95% limit of agreement. A systematic difference in response shift between SF-6D and EQ-5D was also detected in the plot, demonstrated by the presence of a linear relationship between EQ-5D and SF-6D scores. The mean difference between the two measures (EQ-5D minus SF-6D) was 0.43 [SD: 0.40, 95% CI: (-0.36, 1.22)], suggesting that response shift detected by EQ-5D was generally larger than that detected by the SF-6D.

Figure 7.1:



7.4 Discussion

In this 18-month follow-up study of patients undergoing TKR, we detected the presence of response shift and quantified its impact on both HRQoL and utility assessment using the then-test approach. We found that response shift was present at both baseline and six months after TKR, and significantly influenced HRQoL scores. This may have an impact on the use of conventional pre- and post-test methods to assess improvement in HRQoL and utility scores in longitudinal studies. It also suggests that treatment effect may be masked by response shift, due to gradual adaptation to an improving health status of patients (Timmerman et al., 2003; Ring et al., 2005). To the best of our knowledge, this was one of the very few pioneering studies to explore impact of response shift in utility assessment of disease management programs. Besides, it was also the first study exploring response shift among TKR patients using generic measures and over a prolonged 18-month postoperative period. Given that there is great paucity in the published literature on impact of response shift in utility assessment and given that TKR is a commonly performed procedure to manage end-stage osteoporosis, our findings are important in several ways as detailed in the following paragraphs.

First, our results further suggested that response shift did present in the longitudinal HRQoL and utility assessment, especially when a major intervention occurred in between. Though measured with the Western Ontario and Mc Master Universities Osteoarthritis questionnaire (WOMAC) and SF-36, the magnitude and direction of response shift (rating at then-test being worse than the pre-test) at baseline as reported by Razmjou et al. (2006, 2009) in patients undergoing Total knee

arthroplasty were similar to our results. The authors also suggested that the response shift phenomenon can obscure the measurement of the effectiveness and potentially the cost-effectiveness of the surgical intervention. An earlier exploratory study by Bernhard et al., (2001) also detected similar direction of response shift affected utility evaluations of the same health states by cancer patients who undergone chemotherapy or surgical interventions.

Our study further raised concerns over the accuracy of generating utility differences in the conventional post- and pre-test manner to evaluate the effectiveness or cost-effectiveness of an intervention or to make comparisons between several interventions. The substantial changes caused by response shift may also have clinically important implications on drug subsidy or technology assessment, as illustrated below.

In addressing the clinical implications of response shift, the minimal important difference (MID) for a HRQoL or utility score needs to be considered. The MID is defined as the smallest difference in score which patients perceive as beneficial (Jaeschke et al., 1989). The MID of the SF-6D and EQ-5D have been reported as 0.041 and 0.074, respectively (Walters and Brazier, 2005). Based on our results, the quantum of response shift as a percentage of MID changed substantially over time [response shift at baseline: 341% by SF-6D and 97% by EQ-5D; response shift at six months after TKR: 122% by SF-6D]. With regards to economic impact, after adjustment for the response shift observed in this study, the cost-utility ratio would be decreased by almost two-fold, suggesting a substantial increase in the cost-

effectiveness of TKR. For example, when a hypothetical value of USD 10,000 is assigned to account for direct and indirect costs of TKR over 18 months, the impact of response shift in SF-6D scores on cost-utility ratio is as high as USD 29,167/QALY, changing the unadjusted ratio of USD 62,500/QALY, which would be considered not cost-effective (based on a commonly used cut-off point of USD 50,000/QALY) to USD 33,333/QALY, which would be considered cost-effective (Brauer et al., 2005).

Therefore, the impact of response shift could potentially topple the decisions on the approval and subsidy of interventions, especially for cases whose cost-utility ratios are close to the cut-off point before adjusting for response shift. Furthermore, the presence and impact of response shift could also potentially influence comparisons across different studies, if response shift were not fully assessed and adjusted for. Interestingly, although we have shown that response shift generated from SF-6D and EQ-5D differed systematically, it is of note that the impact of response shift on cost-utility ratio in the above example if measured by the EQ-5D would be USD 26,511 /QALY, close to that of the SF-6D.

Second, in terms of contribution to the TKR outcomes research, our data complemented and extended the findings of that the previously reported six-month longitudinal study on another group of patients undergoing TKR using WOMAC, further supporting the idea of incorporating measurement of response shift to more accurately measure treatment effects (Razmjou et al., 2006).

Third, by extending the study period beyond six months, we had the opportunity

to further characterize the nature of response shift and its influence on health status both at baseline and six months after TKR. We found that during the first six-month recovery period, patients may have experienced a larger degree of response shift due to a comparatively faster pace of recovery; subsequently, a plateau in health status could have been reached and maintained, leading to much lesser degree of response shift. Such evidence was also supported by clinical impressions and comments from most of the patients during the telephone survey, who generally expressed a view that their health status had not improved substantially between six and 18 months after TKR (Ethgen et al., 2004). The differing magnitude of response shift at baseline and six months also suggests that recall bias was not a major factor influencing the results of this study, given that a similar magnitude of response shift would be expected if recall bias was present.

Four, the interview during telephone survey at 18 months also provided some other important information for study design and interpretation of results. The discrepancies between quantitative and qualitative assessment of response shift further suggested that patients experienced response shift in an unconscious manner. Therefore, it is necessary to determine both pre- and then-test scores for reference purposes. It was also found that SF-6D was preferred over EQ-5D by 51 patients to assess health status, as the three-level EQ-5D items were felt by subjects to be less accurate a description of their health state when compared to the SF-6D, which had five to six levels for various items. Such information could partially explain the discrepancies between the two measures in detecting response shift at baseline.

Last but not least, the moderate agreement between the SF-6D and EQ-5D in measuring response shift suggested that it was feasible to detect response shift with both measures. However, the systematic difference between the two measures raises a note of caution regarding accurate quantification of response shift in HRQoL and utility assessment.

We also recognize some limitations of the current study. First, the small sample size (N=74) prevented us from generalizing our results to all patients undergoing TKR in Singapore. Further studies by extending the purposive sampling to more cohorts or more TKR centers in Singapore are suggested to confirm the findings from this study. For the same reason, the factors identified in MLR analyses should be considered as exploratory rather than confirmatory. Second, as the EQ-5D pre-test data at six months were not available, we were not able to study the agreement between two measures for this time point to provide more robust results.

7.5 Conclusion

In conclusion, response shift was present and impacted HRQoL and utility assessment among patients undergoing TKR both just prior to and six months after surgery. This suggests that HRQoL and utility evaluations should be performed bearing in mind of potential changes in patients' internal standards which lead to response shift.

Chapter Eight

Applying the Expectancy-Value Model to Understand Health preferences

8.1 Introduction

Since last decade, “health preferences” has become a buzzword due to its fundamental role in developing several widely used preference-based Health-related Quality of Life (HRQoL) measures such as EQ-5D and SF-6D. More significantly, the utility scores generated from health preferences have been incorporated in cost-utility analysis for decision-making in health care resource allocations (Petitti, 1994). Studies have shown that health preferences elicited from the same instrument varied across different health states and populations (Dion et al., 2002; Koh et al., 2003; Shah et al., 2004; Cardarelli et al., 2006; Winkelmayer et al., 2006). Yet, no study has explored the underlying factors systematically, perhaps it is because researchers tried to explain differences from demographic, medical or socio-cultural perspectives in a post-hoc way (Franic and Pathak, 2003; Wildi et al., 2004; Byrne et al., 2004; Chan et al., 2006). Hence, it is of great significance to study health preferences in a prospective manner to elicit a framework of candidate factors. With such structure in mind, both researchers and decision-makers could interpret utility-score based results more rationally.

Actually, if traced back to its origin, health preference was categorized as an attitude in health psychology (Eagly and Chaiken, 1993). Accordingly, “health preference” should be defined as people’s attitude towards a particular health state in terms of satisfaction, distress or desirability (Eagly and Chaiken, 1993; Glaz et al., 2002). In real world practices, “health preferences” are more widely studied, because each respondent is often asked about attitudes on more than one health state or treatment scenario for setting benchmark scores or generating treatment

priorities (Winkelmayer et al., 2006; King et al., 2003).

The most established model to predict attitude in a formulated way is the expectancy-value model (EVM), which provides a popular framework for describing how beliefs are combined to form attitudes (Eagly and Chaiken, 1993). The model proposes that an attitude (interpreted as the evaluation of an attitude object) is a function of the sum of the expected values of the attributes ascribed to the attitude object. The expectancy associated with an attribute is one's subjective probability that the attitude object has the attribute, and the value of an attribute is one's evaluation of it. The expectancy and value associated with each attribute are multiplied together, and these products are then summed to evaluate the overall attitude towards that health state (Fishbein and Raven, 1962; Fishbein and Ajzen, 1975). Yet, as attitudes do not derive exclusively from beliefs that people hold about attitude objects, external variables (like demographic variables, personal traits and/or other variables that are not in the initial EVM) have been introduced to compensate for its inadequacy (Eagly and Chaiken, 1993). Based on aforementioned theory and systematic structure of EVM, we aimed to explore its usefulness in explaining variances in health preference(s) and investigate factors that may influence health preference(s).

8.2 Methods

8.2.1 Study design and subjects

The study was conducted in two phases. Within the first phase, preliminary interviews were carried out to select altogether 5 health states to indicate best health,

worst health, and health with minor or moderate or major problem(s). These 5 health states were aimed to be used as examples to generate health preferences at different levels, rather than as exact representative of each severity level. Hence, final decision of the 5 health states was based on agreement among the majority of respondents and logical concerns of the health states. Due to the wide application of EQ-5D in utility studies and its relative simplicity, it was used as the reference to describe 5 health states. Accordingly, each of the health state incorporates 5 dimensions of mobility, self-care, usual activities, pain/discomfort and anxiety/depression, with a level to indicate no problem (coded as 1), some problems (coded as 2) and extreme problems (coded as 3) in that particular dimension (EuroQol Group, 1990; Dolan, 1997).

After finalization of the selection, focus group discussions were then carried out on the same subjects to identify major and common attitudinal attributes towards those 5 health states. The main reason for excluding other insignificant attitudinal attributes was to reduce response burden, as each additional attribute would generate 10 more questions (5 sets of questions for "expectancy" and "value" for 5 health states).

Eligible participants of the first phase were English-speaking Singaporeans across the three major local ethnic groups (namely, Chinese, Malay and Indian) and aged between 16 and 65 years old. In accordance with rules of thumb, a sample of 3 to 4 groups with 7 people each was planned initially. The final number of groups was determined at the point when information elicited from various groups reached

a “saturation point”, that is, no more new information could be obtained (Holly, 1999; Fern, 2001). Respondents were asked to discuss about any potential attitudinal attributes that may influence health preferences towards the 5 given health states. Besides, relative impacts of elicited attitudinal attributes were ranked at the end of each focus group discussion to identify the most common and important attitudinal attributes. Discussions were audio-recorded and transcribed for content analysis using ATLAS.ti 5.0 Demo (ATLAS.ti Scientific Software Development GmBh, Berlin, 2003-2006). Based on such results, major and common attitudinal attributes were incorporated and developed for EVM questions accordingly. The survey questionnaire was comprised of Visual Analogue Scales (VAS) to measure health preferences (VAS) and EVM items (including external variables and expectancy-value questions of attitudinal attributes). The questionnaire was then assessed for face validity by another 10 respondents. After any necessary amendments, questionnaire was finalized for use in survey.

At the second phase, the self-administered questionnaire was distributed by convenience sampling to eligible respondents, who should be over 16 years old and able to complete the English questionnaire without any assistance. Four trained research assistants were assigned to approach and recruit eligible respondents. Results obtained at this stage were used to explore usefulness of EVM in explaining health preferences and identifying underlying factors.

8.2.2 Measures

Health preferences for each EQ-5D health state were measured with a 0-1 VAS,

of which 0 represents the worst imaginable state and 1 represents the best imaginable state. Appendix 8.1 shows an example of a complete set of questions and answers for both "expectancy" and "value" of one health state. The same set of questions and answers were used to study both "expectancy" and "value" of all the other health states selected in the study. For each of the question, 7-point bipolar Likert scales (range from -3 to +3) were applied to measure "expectancy" [from "-3" (extremely unlikely) to "+3" (extremely likely)] and "value" [from "-3" (extremely bad) to "+3" (extremely good)] respectively. Such bipolar numbering system and the 7-point Likert scales have the advantage of capturing bipolar answers of various respondents in a wide range, compared with monopolar numbering system and 5-point Likert scales.

External variables included demographic information (age, gender, religion, housing, education level, working status, and marital status) and health status (acute disease occurrence in the past month, current chronic disease status and EQ-5D index for health status of the day) (Kind, 1996). Besides, potential psychological influence by others suffering from severe diseases was also incorporated and measured with a 0-10 Likert scale (from "not influenced at all" to "extremely influenced).

8.2.3 Statistical analysis

Health preferences of 5 health states were analyzed both in combination and separately. Student's t-test, analysis of variance (ANOVA) and Post Hoc Tukey test, where applicable, were applied to explore potential external variables that caused

significant differences in health preferences, which would be further included into EVM. Bivariate correlation analysis was performed to study the correlation between health preferences and external variables in the EVM. Pearson partial correlation coefficients were calculated to investigate the relationship between health preferences with the sum of attitudinal attributes and individual attribute respectively, when all the external variables in the EVM were controlled. Cohen's criteria were adopted as a reference for the magnitude of the correlations. Thus, a correlation coefficient of 0.10-0.29 is considered weak, 0.30-0.49 moderate and 0.50 and above as strong (Van et al., 2002).

For each of the 5 health state, multiple linear regression (MLR) analysis was used to explore the explanatory power of EVM for health preferences by examining the sum of attitudinal attributes and external variables separately or in combination. Besides, MLR was further used to examine the explanatory power of each attitudinal attribute in EVM for health preferences as well. When the explanatory power of EVM for health preferences was studied across 5 health states in a mixed scenario, an ordinary least squares (OLS) model was used to account for intra-person correlations. It was because that there might be potential correlations among the health preferences elicited from a same person in such scenario. Adjusted R squares were reported for both of the MLRs and OLS to indicate the variances explained by EVM in health preferences in different scenarios. For all data analysis, p value showing statistical significance was set at 0.05.

8.3 Results

8.3.1 Selection of the 5 health states

All of the 28 respondents agreed to use “11111” and “33333” (see the following paragraph for detailed description of the health states) to indicate full health and worst imaginable health. “Moderate pain or discomfort” was agreed by most of the participants (n=21) to represent minor problem. Similarly, the choice of “21221” and “32323” was also based on agreement among the majority of the respondents and logical concerns to describe rationale health states.

The detailed description of the 5 health states were as follows: (1) Full health: coded as 11111, indicating no problems in mobility, self-care, usual activities, no pain/discomfort, and no anxiety/depression; (2) Minor problem in health: coded as 11121, indicating no problems in mobility, self-care, usual activities, no anxiety/depression, yet moderate pain/discomfort; (3) Moderate problems in health: coded as 21221, indicating no problems in self-care and no anxiety/depression, yet some problems in mobility, usual activities and moderate pain/discomfort; (4) Major problems in health: coded as 32323, indicating some problems in self-care, moderate pain/discomfort, extreme problems in mobility, usual activities and anxiety/depression; (5) Worst possible health: extreme problems in mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.

8.3.2 Generation of attitudinal attributes

Four focus groups with a total number of 28 participants were shown to be adequate for generating attitudinal attributes. Based on the content analysis,

altogether 10 attitudinal attributes were identified, namely, "worsening your quality of life in terms of health", "adding a burden to your family", "making you less independent", "making you unable to work or study", "making you depressed", "making you unable to enjoy social life", "making you unable to have a close relationship with family or friends", "making you losing dignity", "making you less confident" and "making you unable to support your family". Based on the frequency and importance of these attributes as mentioned by the focus group respondents, four following attitudinal attributes were identified as the most important and common ones to influence health preferences of the 5 given health states, namely, “worsening your quality of life in terms of health” (WQoL) (N=28), “adding a burden to your family” (BTF) (N=25), “making you less independent” (MLI) (N=20) and “making you unable to work or study” (UWS) (N=20).

8.3.3 Characteristics and health preferences of survey respondents

Table 8.1 shows health preferences and characteristics of the 232 eligible respondents. ANOVA demonstrated that health preferences varied significantly among the 5 states, ranging from 0.12 for “33333” (worst possible health) to 0.97 for “11111” (full health). Besides, it was found that respondents who are not religious had significantly higher preferences [N=51, VAS score=0.77 (0.17), $p<0.05$] for health state “11121” (minor problem in health) than religious respondents [N=181, VAS score=0.68 (0.22), $p<0.05$]. As for health state “21221” (moderate problem in health), health preferences of those with over 6 years of education [N=218, VAS score=0.47 (0.17), $p<0.05$] were significantly higher than those with 6 or less years of education [N=12, VAS score=0.35 (0.19), $p<0.05$].

Furthermore, respondents who were not working elicited higher preferences on "32323" (major problem in health) [N=18, VAS score=0.40 (0.27), $p<0.05$], compared with the working population [N=214, VAS score=0.25 (0.17), $p<0.05$]. Apart from these three observations, no other significant differences were found.

Table 8.1. Characteristics and health preferences of 232 survey respondents

	N (%) unless stated
Age [Mean (SD)]	27.7 (15.07)
Female	114 (49.1)
Ethnicity	
Chinese	153 (65.9)
Malay	36 (15.5)
Indian	40 (17.2)
Years of Education	
≤ 6 years	12 (5.2)
7-12 years	165 (71.1)
≥ 13 years	54 (23.3)
Housing	
Public housing	189 (81.5)
Private housing	39 (16.8)
Working	214 (92.2)
Married	73 (31.5)
Religious *	181 (78.0)
Presence of acute medical conditions in the past month †	157 (67.7)
Presence of chronic medical conditions ‡	87 (37.5)
Past experience with people suffering from severe diseases	132 (57.6)
Potential psychological influence by others suffering from severe diseases §[Mean (SD)]	7.43 (2.66)
EQ-5D index [Mean (SD)]	0.96 (0.55)
Health preferences [Mean (SD)]	
11111	0.97 (0.09)
11121	0.70 (0.21)
21221	0.48 (0.18)
32323	0.26 (0.19)
33333	0.12 (0.21)

*: Religions included Buddhism, Taoism (Chinese traditional beliefs), Islam, Hinduism, Sikhism, Catholicism, Christianity and others etc.
†: Acute medical conditions included running a nose, sore throat or cough, vomiting or diarrhea, headache lasting more than one day, sleeping problems and body injuries.
‡: Chronic medical conditions included diabetes, hypertension, heart disease, stroke, asthma or other lung disease, rheumatism, back pain or other bone or muscle illness, mental illness and kidney problems on dialysis.
§: potential psychological influence by others suffering from severe diseases was also incorporated and measured with a 0-10 Likert scale (from "not influenced at all" to "extremely influenced).

8.3.4 Correlation between various components of EVM and health preferences

In the bivariate analysis, none of the external variables in the EVM was shown to be significantly correlated with health preferences. Table 8.2 shows partial correlation coefficients between health preference(s) and attitudinal attributes in terms of the sum of attitudinal attributes and individual attributes, when all the external variables in the EVM are controlled.

When data were analyzed across all 5 health states with control over all external variables and subjects, correlation between the sum of attitudinal attributes and health preferences was as strong as 0.78 ($p<0.01$). Besides, all four attitudinal attributes demonstrated strong correlation with health preferences, ranging from 0.68 to 0.73 ($p<0.01$). Comparatively, when further control over health state was added to remove the impact of severity level, correlation coefficients between attitudinal attributes and health preferences ranged from 0.25 to 0.35 ($p<0.01$). As for the data analysis of each individual health state, except for "MLI" in health state "11111", significant correlations between attitudinal attributes and health preferences were found with correlation coefficients ranging from 0.14 to 0.42 ($p<0.01$).

Furthermore, a trend was observed. As severity of health state increased, scores of attitudinal attributes (shown in Table 8.3) and health preferences decreased, suggesting less positive attitudes.

Table 8.2 Correlation between health values and attitudinal attributes*

Health value(s) by health state(s)	Attitudinal attributes				
	SUM	WQoL	BTF	MLI	UWS
5 health states mixed (1)†	0.78¶	0.69¶	0.71¶	0.68¶	0.73¶
5 health states mixed (2)‡	0.35¶	0.28¶	0.25¶	0.28¶	0.29¶
By individual health state					
11111	0.27¶	0.31¶	0.15¶	0.10¶	0.25¶
11121	0.42¶	0.26¶	0.27¶	0.37¶	0.30¶
21221	0.31¶	0.26¶	0.25¶	0.21¶	0.28¶
32323	0.31¶	0.26¶	0.25¶	0.21¶	0.28¶
33333	0.20¶	0.20¶	0.16¶	0.17¶	0.14¶

*The following external variables were controlled in the partial correlation analysis: age, sex, religion, housing, education level, working status, and marital status, and health status (presence of acute disease occurrence in the past month, presence of current chronic disease status, and EuroQol Five-Dimension (EQ-5D) index for health status of the day) and potential psychological influence by others suffering from severe diseases.

†With additional control over subjects.

‡With additional control over subjects and health states.

¶$P < 0.01$.

BTF, adding a burden to your family; MLI, making you less independent; SUM, sum of products of the four additional attributes as WQoL, BTF, MLI, and UWS;UWS, making you unable to work or study; WQoL, worsening your quality of life in terms of health.

Table 8.3 Score distribution of attitudinal attributes of Expectancy-Value Model by health state

EQ-5D Health State	SUM	WQoL	BTF	MLI	UWS
11111	18.15	4.61	4.54	4.01	4.99
	(11.85)	(3.87)	(4.53)	(4.01)	(3.63)
11121	2.32	-0.22	0.24	0.83	1.46
	(10.16)	(2.80)	(3.36)	(3.29)	(3.36)
21221	-8.83	-2.53	-2.37	-2.31	-1.62
	(10.71)	(3.19)	(3.19)	(3.01)	(3.27)
32323	-20.36	-5.17	-5.49	-5.03	-4.66
	(11.95)	(3.65)	(3.58)	(3.74)	(3.74)
33333	-23.27	-5.70	-6.79	-5.28	-5.50
	(12.27)	(4.56)	(3.79)	(4.52)	(4.00)

Data were shown as Mean (SD)
SUM: sum of products of the four additional attributes as WQoL, BTF, MLI and UWS
WQoL: worsening your quality of life in terms of health
BTF: adding a burden to your family
MLI: making you less independent
UWS: making you unable to work or study

8.3.5 Explanatory power of EVM

When 5 health states were mixed, EVM explained up to 62% of the variances in health preferences. When EVM was applied to each individual health state, the explanatory power of EVM was reduced to a range between 8% and 23%. Despite such, results showed that compared with models only incorporating external variables, EVMs had much higher and significant explanatory powers (Shown as Table 8.4).

Table 8.5 presents the contribution of each attitudinal attributes to explaining health preference(s), when all the external variables are controlled. As for the mixed health states, UWS, BTF and WQoL had larger influence on health preference than MLI. “WQoL” had significant and dominant impacts on preferences towards

"11111", "11121" and "21221". "UWS" and "BTF" were the two major influential attributes on preferences of "32323", compared with the other 2 attributes. As for the worst state "33333", "MLI" generated significantly larger impact on health preferences.

Table 8.4: Comparison of explanatory power of expectancy-value model versus external variables only[*]

Health State	Regression model (N=232)	
	expectancy-value model [†]	External variables only
5 Health states mixed	0.62[‡]	-0.004
By Health State		
11111	0.08 [‡]	0.02
11121	0.15 [‡]	0.05[§]
21221	0.08 [‡]	0.02
32323	0.13 [‡]	0.04
33333	0.23 [‡]	0.03

[*]: Data presented as adjusted R square
[†]: Expectancy-Value Model included the sum of attitudinal attributes and external variables (including age, gender, religion, housing, education level, working status, and marital status, acute disease occurrence in the past month, current chronic disease status and EQ-5D index for health status of the day and potential psychological influence by others suffering from severe diseases). For the mixed scenario of 5 health states, ordinary least squares regression was used to account for intra-person correlation.
[‡]: $p<0.01$
[§]: $p<0.05$

Table 8.5: Contribution of each attitudinal attribute to explaining health preferences [*]

Health State	Attitudinal attributes of the Expectancy-Value Model (N=232) [†]			
	WQoL	BTF	MLI	UWS
5 health states mixed	0.22 **	0.24 **	0.13 **	0.28 **
By Health State				
11111	0.28 **	0.01	-0.06	0.14
11121	0.15 **	0.07	0.04	0.14
21221	0.12	0.05	0.01	0.08
32323	0.09	0.14	0.04	0.15
33333	0.09 **	0.12	0.27 **	0.14

*: Standardized coefficients (beta)
[†]:The following external variables were controlled in the multiple liner regression: age, sex, religion, housing, education level, working status, and marital status, and health status (presence of acute disease occurrence in the past month, presence of current chronic disease status, and EuroQol Five-Dimension (EQ-5D) index for health status of the day) and potential psychological influence by others suffering from severe diseases.
[‡]: $P < 0.01$.
BTF, adding a burden to your family; MLI, making you less independent; UWS, making you unable to work or study; WQoL, worsening your quality of life in terms of health.

8.4 Discussions

In this exploratory study to investigate the power of EVM in explaining health preferences, we found that compared with models that only incorporated external variables, EVM that incorporated both external variables and attitudinal attributes could explain a much larger proportion of the variances in health preferences. Besides, EVM could further explore the contributions of each component, either attitudinal attribute or external variable, to identify important factors that influence results of health preferences and quantify their magnitude by regression analyses. In our study, it was interesting to note that the magnitudes of influence of each individual attitudinal attribute differed across various health states. Such differences were probably related to the differences in health states, which would trigger

respondents' different expectancy towards the attitudinal attribute.

Our results provide important implications in understanding health preferences in several ways:

Basically, EVM-based utility measure could be used to help explain and understand the differences in utility values elicited by patients themselves from the same utility measure. Cautions should be taken not to apply EVM to study the results generated from different-based utility measures (e.g. time trade-off, standard gamble, VAS, etc), as these measures were reported to elicit different scores due to methodology concerns (Krabbe et al., 1997). We suggested using EVM as a complementary tool when the interest was to understand the differences in preference-based utility values across different studies or to check why the utility values elicited from patients were different from expectations of health care practitioners or decision makers during the process of assessing disease management programs.

The promising explanatory power of EVM for health preferences could also potentially help evaluating disease management programs in other ways. It could be applied by health care practitioners and decision makers to better understand patients' psychological concerns about the treatment or medication. Based on such information, they could partially figure out why certain patients refuse to accept or poorly comply with the intervention. For example, Polsky et al. (2002) found that a woman's preference of surgical treatment for breast cancer was largely influenced by her perception of its short-term benefits rather than long-term ones. Hence, even though the surgical treatment would be a long-term cost-effective measure to

improve patients' treatment outcomes and health-related quality of life, the misconceptions from patients might even eventually alter the fact and consequently the choice (Polsky et al., 2003). Therefore, it would be useful to put such perception into EVM as one of the attitudinal attributes. By doing so, decision makers and practitioners could have identified underlying reasons why the elicited utilities do not match with their assumptions, which might topple the sequential judgment on the cost-effectiveness of the intervention.

Furthermore, on individual basis, health care practitioners could help patients rectify certain wrong perceptions of attitudinal attributes towards treatment, so as to improve patient-practitioner communication, treatment compliance which might contribute to better treatment outcomes and even the cost-effectiveness of the treatment. Such application could also be expanded to other disease management programs, such as health care education on attitudes towards smoking cessation, HIV prevention, etc. In such cases, EVM could be used as part of the effectiveness assessment to evaluate changes in attitudes.

However, limitations together with suggestions for future studies should be noted in this exploratory study as well:

First, in general, EVM tends to be less powerful to explain variances in health preference of individual health state compared with mixed states. A potential contributor of such poor performance for individual health states might be due to the limitation of Likert Scale to measure attitudinal attributes accurately. The 7-

point Likert scale only allowed respondents to rate “expectancy” and “value” with 1 out of the 7 numbers. However, health preferences were elicited on a continuous VAS from “0-1”, which provided infinite choices. Such incompatibility in the scaling method could have caused poor differentiation power of EVM when variances of health preferences in each state were rather small. Hence, we suggested future studies with compatible continuous scales to measure attitudinal attributes be carried out first for validation purposes. If validated, further research could be focused on the explanatory power of EVM constructed on these new scales.

Second, the attitudinal attributes generated in the current study may not be applicable to other health scenarios or populations due to its potential specificity and sociodemographic influences. Hence, it is suggested that the attitudinal attributes of the current study should be validated first before its application in a different population or a new health scenario. If the validation is failed, exploration of the applicable attitudinal attributes should be generated from the scratch. Although it would be clearer to have common attributes for comparison studies, those specific additional attribute(s) could provide additional useful information to explain significant variances.

Third, the convenience sampling used in this study may potentially hamper the generalization of the study results to the general population in Singapore. Further studies using proportional quota sampling are suggested to confirm our study findings and also to improve the generalizability of the study results.

8.5 Conclusions

In summary, our results showed that EVM was useful in explaining variances of health preferences. However, its power to predict small variances might be restricted due to limitations of current Likert Scale to measure “expectancy” and “value” in EVM. With further improvement and validation of a compatible continuous scale for more accurate measurement, EVM is expected to explain health preferences to a larger extent. Future studies are suggested to explore the power of EVM in explaining health preferences towards different health or treatment scenarios to further demonstrate its robustness and wider adaptabilities.

Chapter Nine

Development and Validation of a Scale to measure Patients' Trust in Pharmacists

9.1 Introduction

In the current era of disease management, with the increasing emphasis on pharmaceutical care, the roles of the pharmacists have been undergoing vigorous expansion. Studies have shown that better pharmaceutical care contributed to less drug-related morbidity and mortality, improved clinical outcomes and health-related quality of life (HRQoL), and lowered medical costs, which may further contribute to the cost-effectiveness of disease management programs (Penna, 1990; Benny et al., 2000; Hermansen and Wilderholt, 2001).

A key success factor of pharmaceutical care is the quality of patient-pharmacist relationship. In such relationship, patients grant authority to pharmacists to manage their health and well-being. In turn, pharmacists accept responsibility and do good to patients (Penna, 1990; Helper and Strand, 1990). Because of the vulnerability of patients and uncertainties of outcomes, patient-pharmacist relationship is largely influenced by the level of patients' trust in pharmacists (Moorman et al., 1993). From patients' perspective, trust in pharmacists could be defined as "patients' willingness to be vulnerable to the actions of pharmacists based on the expectation that pharmacists will do what is best for patients, irrespective of patients' ability to monitor pharmacists." (Mayer et al., 1995).

Based on the literature search in PubMed (1966 to Oct 2007), there was no published scale to measure patients' trust in pharmacists. In order to assess influence of trust on disease management related outcomes (i.e., effectiveness or cost-effectiveness of the treatment, adherence to drug therapy, satisfaction with the

pharmacy service, quality of pharmaceutical care, etc.) as demonstrated in studies on other medical practitioners, a reliable and valid instrument to test patients' trust in pharmacists should be developed first, which was actually the primary objective of our study (Anderson and Dedrick, 1990; Thom et al., 2004; Piette et al., 2005)

Although practical patterns are different between pharmacists and other healthcare practitioners, items and dimensions in trust scales of other medical practitioners could still provide reference for the scale development. It was found that items in trust scales could be summed into two overarching dimensions as technical competence and benevolence (Thom and Campbell, 1997; Leisen and Hyman, 2001). Technical competence might include evaluating problems thoroughly, providing appropriate and effective treatment, predisposing factors and structural and staffing factors. Benevolence dimension might comprise understanding patients' individual experiences, expressing caring, communicating clearly and completely, building partnership and sharing power, demonstrating honesty and respect, and keeping information confidential (Anderson and Dedrick, 1990; Thom and Campbell, 1997; Kao et al., 1998; Swan et al., 1999; Leisen and Hyman, 2001).

Additionally, to test construct validity, it was hypothesized that patients' trust in pharmacists were positively correlated with "satisfaction with pharmacists' service", "returning for care" and "preference of decision-making patterns" respectively (Anderson and Dedrick, 1990; Thom and Campbell, 1997; Kao et al., 1998; Leisen and Hyman, 2001; Dugan et al., 2005).

9.2 Methods

9.2.1 Study design and subjects

The whole study was divided into two phases: scale development and scale validation.

9.2.1.1 Scale development

A Literature review on trust scale in other medical professions was performed for reference. A focus group approach was then used to determine whether the concept, domains and items based on literature review were relevant and to explore any potential new domains or items. Eligible participants were recruited from the final-year pharmacy undergraduates at the National University of Singapore, who had consulted and obtained medications from pharmacists during the past six months before the study. The rationale to recruit final-year pharmacy undergraduate students as focus group participants was that their valuable experience as intern pharmacists could add in the perspective from pharmacists as well as third-party observers.

In accordance with rules of thumb, a sample of three to four groups with seven people each was planned initially. The exact number of groups was determined by the reach of "saturation point" (Hall et al., 1999). Focus group discussions were audio-recorded and transcribed for content analysis using ATLAS.ti 5.0 Demo (ATLAS.ti Scientific Software Development GmBh, Berlin, 2003-2005).

Based on the results of focus group approach and literature review, the candidate version of "Trust in Pharmacists Scale" was developed. The scale was structured with

both favorably and unfavorably worded items to avoid respondents' blind agreement with statements regardless of the content (Ware, 1978). Answers were formatted with a 7-point Likert scale (with -3=totally disagree and 3= totally agree) to improve score distribution (MacKeigan and Larson, 1989)**.** Then the questionnaire was evaluated by 10 other pharmacy graduate students for face validity (the relevance of items to measure "patients' trust in pharmacists") and clarity of items. An expert panel of three experienced pharmacists was asked to assess content validity, that is, how well items represented the specific intended domains based on experts' judgment (Armstrong, 2005). Pilot testing of the revised version was undertaken by a convenience sample of another 77 final-year pharmacy undergraduates, who had not participated in the focus group discussion. Time of completion, comments on questionnaire were also collected to ensure non-excessive burden on the respondents and to provide further suggestions on item modifications.

9.2.1.2 Scale validation

After necessary amendments, the finalized questionnaire was distributed by research assistants to a sample of 1,200 English-speaking Singaporeans at local neighbourhoods and community centers using purposive sampling. It was designed that study subjects should be across the three major local ethnic groups (Chinese, Malay and Indian) from 3 age groups (18-35 yrs; 36-55 yrs; 56 yrs and above) with an equal ratio among ethnic groups and 2:2:1 ratio among age groups to explore factor structure, reliability and validity of the questionnaire. The uneven ratio of age groups was due to the fact that English-speaking elderly population is small in Singapore and they might have cognitive dysfunction as well (Singapore census of population, 2005).

Eligible respondents should be able to complete English questionnaire without any help and have consulted or obtained medications from a pharmacist during the past six months.

9.2.2 Measures

Besides the trust scale, several other measures were included in the finalized questionnaire: demographic information (age, gender, ethnicity, housing, education level); current chronic disease status, basic information relevant to the scale (time period of last visit to a pharmacist; type of pharmacist visited last time); six-item scale to assess patients' satisfaction with the pharmacists' service; and two items to assess the behavior intents [willingness to return for care on the 7-point Likert scale, preferred decision-making pattern on medication on a 5-point Likert scale (1=totally by myself, 5=totally by pharmacist)] (MacKeigan and Larson, 1989; Davies and Ware, 1991; Kraetschmer et al., 2004; Singapore census of population, 2005).

9.2.3 Statistical analysis

9.2.3.1 Scale development

Item means and standard deviations were assessed to determine whether adequate variability and symmetry in score distribution were achieved in the pilot testing version.

9.2.3.2 Scale validation

Sample characteristics were descriptively analyzed. Response means and standard deviation were calculated to determine variability and symmetry in score distributions,

which indicated discriminatory power. To select candidate items, item analyses were done to ensure the corrected item-total correlation coefficients should be greater than 0.30 for finalized items (Hinton et al., 2004). Exploratory factor analysis was used to determine dimensions of trust. Principal components analysis of the partial correlation matrix was used to identify groups of homogeneous items suitable for measuring each dimension of trust. The number of factors selected for Varimax rotation was determined by a combination of criteria: (1) the roots criterion of selecting factors with eigenvalues to be greater than 1; (2) the Scree test to examine a plot of eigenvalues and stop factoring at the point where the pot begins to level off; and (3) the interpretability and meaningfulness of trial factor rotations. Each factor should have two or more loadings above 0.40 to make a rotated factor interpretable. In addition, the items loading on one factor should fit together logically (Kim and Mueller, 1978; Stewart, 1981; Gorsuch, 1983; Beun, 1996). Tentative scale was then composed of those items with factor loading above 0.40 on one factor and lesser loadings on other factors to represent each dimension. If an item loaded above 0.40 on more than one factor, assignment of the item was to be based on logical fit and verification by item analysis (Stewart, 1981; Gorsuch, 1983; Beun, 1996).

As for reliability, internal consistency was assessed by Cronbach's alpha. Construct validity was studied using Pearson's correlation coefficients to indicate the association between "patients' trust in pharmacists" with "patients' satisfaction with pharmacists' service", "willingness to return for care" and "preferred decision-making pattern on medication".(Beun, 1996)

9.3 Results

9.3.1 Scale development

Altogether four focus groups were shown to be adequate for item generation. Eighteen items were identified in literature review, focus group discussion and study team discussion. (Table 9.1) Because most of the items generated from the focus group discussion were the same as those identified in other studies, relevance of those items were analyzed based on a 5-point Likert scale (1=least relevant, 5=most relevant) (Items 1 to 17 in Table 9.1). Results showed that age, gender, ethnicity of the pharmacist might be less relevant to patients' predisposing trust in pharmacists than the rest of the items. There were two brand new items generated from the focus group discussion: type of the pharmacist and set-up of the pharmacy counter. Some of the participants thought aloud that they preferred hospital pharmacists due to their rich experience with various illnesses. Additionally, "a neat and tidy pharmacy counter" was regarded as a reflection of professionalism and efficiency of pharmacists. Based on literature review, another item on global trust was also included (Item 18 in Table 9.1) (Leisen and Hyman, 2001).

The questionnaire was assessed as having face validity by the pharmacy postgraduate students and as content valid by the expert panel of pharmacists. In the pilot test, the average completion time was 7.4 minutes. Score distribution of 18 items was shown to achieve good variability (SD>1.0) and symmetry (means ranged from -1 to 1). In the open-ended question on readability of the questionnaire, quite a few respondents suggested to avoid using negative worded items so as to minimize confusion. Except for some minor wording changes, no other problems were raised.

After minor revisions, the questionnaire was finalized with 18 items as shown in Table 9.2. (See Appendix 9.1 for the scale)

Table 9.1: Item generated from focus group approach and response analysis

	Item	*Source **	*Relevance* † *Mean (SD)*
1	Demonstrating up-to-date knowledge	3	4.32 (0.72)
2.	Evaluating medical problems thoroughly	3	4.36 (0.68)
3.	Keeping information totally private (confidentiality)	3	4.75 (0.52)
4.	Demonstrating honesty when a mistake is made	3	4.32 (0.77)
5.	Expressing concern	3	4.43 (0.69)
6.	Communicating clearly and completely	3	4.61 (0.50)
7.	Showing sufficient respect	3	4.57 (0.69)
8.	Providing effective medication at a reasonable price	4	4.82 (0.48)
9.	Allowing shared decision-making pattern when there are alternatives.	2	3.89 (0.69)
Predisposing factors			
10.	Age of the pharmacist	3	2.50 (1.28)
11.	Gender of the pharmacist	3	2.25 (1.18)
12.	Ethnicity of the pharmacist	1, 4	2.11 (1.10)
13.	Type of the pharmacist (Hospital, polyclinic, community, etc.)	1	N/A
14.	Past experience with pharmacists	4	4.00 (0.90)
15.	Recommendation by others (friends, neighbors，relatives, etc.)	3	3.68 (0.72)
16.	Set-up of the pharmacy counter	1	N/A
17.	Professional appearance	3	4.14 (0.71)
18.	Global trust in pharmacist	2	N/A

*: 1= generated from focus group only; 2=generated from literature review only; 3=both 1 and 2; 4=generated by the study team only

†: Relevance is scored on a 5-point Likert scale (1=most irrelevant; 2=somewhat irrelevant; 3=neutral; 4=somewhat relevant; 5=most relevant) from 28 participants. It is presented as Mean (SD).

N/A: Not applicable

Table 9.2: Item analyses of 18 candidate items

Candidate Items	Score *	Corrected item-total correlation coefficients
1. I trust a pharmacist who has updated knowledge	1.58 (0.98)	0.39
2. I trust the pharmacist if he/she evaluates my medical problem thoroughly.	1.59 (0.99)	0.45
3. I trust the pharmacist who could keep the information we discuss totally private.	1.63 (1.05)	0.43
4. I trust the pharmacist if he/she tells me about a mistake he/she has made on my medication.	0.95 (1.35)	0.28
5. I trust the pharmacist if he/she expresses concern and talks to me with reassuring and comforting words.	1.47 (0.99)	0.41
6. I trust the pharmacist if he/she could communicate with me clearly and completely.	1.79 (0.87)	0.46
7. I trust a pharmacist who shows sufficient respect for me.	1.75 (0.74)	0.43
8. I trust the pharmacist if he/she provides me with effective medication at a reasonable price to me.	1.60 (1.03)	0.46
9. I trust the pharmacist if he/she allows me to make decision on which medication to take when there are alternatives.	0.99 (1.19)	0.33
10. I trust an older pharmacist more than a younger one.	0.44 (1.42)	0.32
11. I trust a hospital pharmacist more than other types of pharmacist (community, polyclinic pharmacist, etc.)	0.31 (1.41)	0.28
12. I trust a pharmacist with professional appearance.	1.25 (1.07)	0.46
13. I will trust other pharmacists if I have had pleasant past experience with another pharmacist.	0.25 (1.48)	0.14
14. I trust a pharmacist of the same race as myself more than the other races.	-0.51 (1.59)	0.27
15. I trust a pharmacist of the same gender as myself more than the opposite gender.	-0.56 (1.53)	0.19
16. I trust a pharmacist who has been recommended by others (i.e., friends, neighbors or relatives, etc.)	1.13 (1.06)	0.39
17. I trust a pharmacist whose set-up of the counter is neat and tidy.	1.19 (0.95)	0.37
18. I trust a pharmacist so much that I always try to follow his/her advice.	0.84 (1.14)	0.38

*: Score is presented as Mean (SD).

9.3.2 Scale validation

The finalized questionnaires were distributed to 1,206 eligible respondents. Data analysis was based on 1,196 respondents with complete answers on all 18 items of the trust scale. Demographic and background information was summarized in Table 9.3.

Item analyses (Table 9.2) showed that five items (Items 4,11,13,14, and 15) were below the criteria of 0.3 in item-total correlation (Hinton et al., 2004), so they were dropped in the first round. When the 13 candidate items underwent the same analysis again, item 10 was found to be below the criteria (coefficient =0.12 <0.3), leaving 12 items for subsequent analysed.

Table 9.3: Demographic and background information of respondents

	N (%) unless specified otherwise
Age *	38.6 (14.9)
Female	617 (51.6)
Ethnicity	
Chinese	460 (38.5)
Malay	373 (31.2)
Indian	363 (30.4)
Housing	
Public housing	927 (77.5)
Private housing	254 (21.2)
Presence of chronic medical problems	526 (44.0)
Education level	
≤6 yrs of education	137 (11.5)
7 to 10 yrs of education	646 (54.0)
⩾11 yrs of education	394 (32.9)
Last visit to a pharmacist	
Within 1-3 months	671 (56.1)
Within 4-6 months	525 (43.9)
Type of pharmacist visited last time	
Hospital	294 (24.6)
Polyclinic	388 (32.4)
Community	507 (42.4)

*: Age is presented as Mean (SD).

Based on the criteria of eigenvalue and Scree plot, three factors were identified that accounted for 55% of the total variance with the first factor explaining up to 36%. Two items (Items 3 and 5) had dual loadings (loadings greater than 0.40 on two factors). Scale assignments for them were made on the basis of logical fit with other items loading on the two factors under consideration. Table 9.4 lists the factor loading and assignment of the 12 items.

Table 9.4: Factor analysis and reliability of the 12-item trust scale

	Item No.	Factor loading	Corrected item-total correlation	Cronbach's alpha if item deleted
Factor 1 (benevolence)	3	0.42	0.56	0.81
	5	0.46	0.52	0.81
	6	0.68	0.60	0.80
	7	0.79	0.55	0.81
	8	0.73	0.53	0.81
	9	0.53	0.36	0.82
Factor 2 (technical competence)	1	0.84	0.52	0.81
	2	0.82	0.60	0.80
Factor 3 (global trust)	12	0.62	0.43	0.82
	16	0.67	0.35	0.82
	17	0.74	0.43	0.82
	18	0.59	0.38	0.82

Factors were labeled as follows according to the order of extraction: (1) benevolence (including six items: confidentiality, expressing caring, communicating clearly and completely, showing sufficient respect, providing effective medication at a reasonable price and allowing shared decision-making pattern when there are alternatives); (2) technical competence (including two items: demonstrating up-to-date knowledge and evaluating medical problem thoroughly); and (3) global trust including all other aspects that do not exclusively fit in any dimension (including four items: blind trust in pharmacists and predisposing factors such as recommendation by others, set-up of pharmacy counter and professional appearance) (Anderson and

Dedrick, 1990; Thom and Campbell, 1997; Kao et al., 1998; Swan et al., 1999; Leisen and Hayman, 2001).

Due to the observation that the correlation coefficients among three factors via Promax rotation were all less than 0.5, discriminatory power was demonstrated and a three-dimension scale structure was suggested (Kim and Mueller, 1978). As shown in Table 9.4, item-scale correlation coefficients were between 0.35 and 0.60. The Cronbach's alpha was 0.83, indicating high internal consistency (Beun, 1996). Construct validity was demonstrated by the finding that, as hypothesized, patients' trust in pharmacists was positively correlated with patients' satisfaction with pharmacists' service ($r=0.54$, $P<0.001$), returning for care ($r=0.30$, $P<0.001$) and preference of medical decision-making pattern ($r=0.16$, $P<0.001$).

9.4. Discussion

In the current study, we developed and validated a 12-item scale to measure patients' trust in pharmacists. The 12-item scale which demonstrated high reliability and good construct validity could be further categorized into 3 factors, namely, technical competence, benevolence and global trust. To the best of our knowledge, the current study was the first one to develop and validate a scale to measure patients' trust in pharmacists. Our study provided important contributions and implications in several ways:

First, as mentioned in the introduction, pharmacists are not only the key personnel to deliver high-quality pharmaceutical care, but they are expected to

contribute significantly to disease management due to the increasing prevalence of chronic diseases in our rapidly ageing societies and urging request for cost containment. Hence, the relationship between pharmacists and patients might directly or indirectly affect the final results of whether the disease management program such as effectiveness or cost-effectiveness. Therefore, the availability of such scale will enable the incorporation of “patients’ trust in pharmacists” as either a dependent or an independent variable in pharmacoeconomics and pharmaceutical care studies in various healthcare settings. Such evaluations would not only further contribute to the determination of the impact of patients' trust in pharmacists on economic analysis, but also to the better understanding of the key elements in such trust relationship between pharmacists and patients.

Second, the current study served as a good example to show that in the application of cost-utility or other quality of life studies in the Asian countries, sometimes there is a need to develop and validate a new instrument from scratch. Such circumstances might include but are not limited to the universal lack of the measure or measures developed in Western countries could not be readily adapted to the East due to socio-cultural difference for instance. In such case, a systematic approach would be recommended to ensure the robustness of the results. For example, in the current study, items were generated by literature review and focus group approach, followed by item refinement using panel review and pilot testing; and scale validation was conducted by response analysis, reliability and validity tests and exploratory factor analysis.

Third, interestingly, the scale to measure patients' trust in pharmacists was shown to have similar dimensionalities to the scale to measure patients' trust in physicians. The two common dimensionalities were technical competence and benevolence. This implied that patients might tend to trust health care professionals based on similar criteria. In other words, technical competence and benevolence were the two most sought-after qualities that would help to build up patients' trust in health care professionals. If so, it is worthwhile for healthcare providers to put special efforts in enhancing technical competence and benevolence to obtain more trust from patients. A series of desired outcomes might be consequentially achieved, such as better communication between patients and health care professionals, enhanced compliance, which might translate into improved clinical outcomes, better health-related quality of life, and even positively influence the cost-effectiveness of disease management programs (Beun, 1996).

In the mean time, several limitations together with suggestions for future studies should also be noted:

First, since patients' trust is a changeable psychological trait and not a steady state (Bruce, 2001; Thom et al., 2004), retest was not carried out in our study. Therefore, test-retest reliability remains unknown for the scale, retest may need to be carried out in future studies to explore the feasibility of retest and verify the robustness of the scale.

Second, despite potential advantages of thinking of trust in perspectives of both patients and pharmacists, final-year pharmacy undergraduates may not well represent

pharmacy-visiting patients in the real world. Hence, it is possible that the items generated from those undergraduates might be different from patients with diversified backgrounds.

Third, respondents of the current study tended to be healthier and better educated compared with the average pharmacy-visiting patients, due to use of self-administration purposive sampling methodology. Such potential bias may lead to different emphasis on the items of trust in pharmacists, which may lead to somewhat different chosen items. Hence, it is suggested that future study be sampled on a more general heterogeneous population to further validate the scale.

9.5 Conclusion

The 12-item scale, constructed on three factors, namely, benevolence, technical competence and global trust, to measure patients' trust in pharmacists demonstrated high reliability and good construct validity. The availability of such scale would enable the incorporation of "patients' trust in pharmacists" as an independent variable in pharmacoeconomic evaluations such as CUA to assess disease management programs in various health care settings.

Chapter Ten

A Recapitulation of Major Findings, Contributions, Limitations, and Future Studies

10.1 General Introduction

In this last chapter, I would like to take the opportunity to recapitulate those major findings made in this thesis and evaluate how adequate have the studies performed answered the original questions posed at the beginning of the thesis. At the same time, the major contributions and limitations of our studies would also be highlighted, followed by some suggestions of future studies that would further contribute to the application of CUA in assessing disease management in Asian countries.

10.2 Major findings

In the section of research questions in the Introduction Chapter (Section 1.5), we started off with the following questions in mind:

(1) Is there any linguistic or cultural barrier in the adaptation of an English HRQoL or utility instrument into a non-English version?

(2) What are the factors that have been found to influence the application of CUA in the published literatures so far?

(3) Are there any other new factors that may potentially influence the application of CUA of disease management programs in Asia?

Specific to the three broad questions as mentioned above, individual studies were conducted as presented from Chapter 2 to Chapter 9. Major findings from these studies to answer each of the 3 research questions are highlighted in the following sections.

10.2.1 Addressing Research Question 1- impact of linguistic or cultural barrier

In Chapter 2, using the translating and adapting the English Audit of Diabetes-Dependent Quality of life (ADDQoL) into a Chinese version as an illustrative example, we detected that minor, potential linguistic and cultural differences did exist between the original English version and the translated Chinese version. More importantly, to conquer such barriers, we found that the universalist approach was essential for the systematic translation and adaptation process. Such approach ensured that all important equivalences, namely, conceptual, item, semantic, operational, measurement and functional equivalences were investigated in a sequential manner and potential issues were picked up and solved throughout the whole process. In the end, we demonstrated that all those six types of equivalences were achieved and the Chinese ADDQoL was a reliable and valid diabetes-specific HRQoL instrument to be used in future studies among Chinese-speaking diabetic patients in Singapore.

10.2.2 Addressing Research Question 2- Factors identified from published literature

In Chapter 3, by conducting a qualitative literature review to find out factors that have been demonstrated to influence CUA application or CUA results, altogether 20 factors were identified in the published literature. They could be further grouped into five categories as treatment-related (duration, efficacy, and frequency of treatment), disease-related (severity level, risk level, incidence rate, prevalence rate, disease progression rate and survival length), patient-related (age, gender, race/ethnicity and compliance), cost-related (treatment cost, incorporation of indirect cost), and methodology-related factors (discount rate, QALY elicitation method, statistical

uncertainty, handling of confounding variables and reliability of data source). We also found that most of the CUA studies included in the literature review were conducted in Western countries, suggesting that very few studies in Asia have adequately explored the potential influence of those factors on their CUA results. It further encouraged us to devote our efforts to enhance the robustness of applying CUA studies in Asian countries.

10.2.3 Addressing Research Question 3-Exploring other new factors

In the current thesis, we focused our efforts on the exploration of two patient empowerment-related factors (namely, disease knowledge and health literacy) and three health psychology-related factors (namely, response shift, expectancy-value, and trust in pharmacists).

10.2.3.1 Patient-empowerment factors

In chapter 4, we used diabetes knowledge as an illustrative example to explore its potential impact on health utility values. In this preliminary study among English-speaking diabetic patients, we found that the correlation between diabetes knowledge (measured by GDKT) and health utility values (measured by generic health utility instruments as EQ-5D and SF-6D) was rather minimal, suggesting the impact of diabetes knowledge on the generic health utility values was lacking. Although not statistically significant, diabetes knowledge was found to be weakly correlated with disease-specific HRQoL score as measured by ADDQoL. However, after adjusting for other variables, diabetes knowledge had no further impact on ADDQoL scores, suggesting potential lack of impact on health utilities generated by disease-specific

health profile measures.

With special regards to health literacy, in Chapter 5 we found that in the cross-sectional study among patients with rheumatic diseases, health literacy level as measured by the word recognition test, called the "Rapid Estimate of Adult Literacy in Medicine" (REALM) did not impact health utility scores as measured by either SF-6D or EQ-5D. Furthermore, it was also found that health literacy level did not influence HRQoL as measured by SF-36 in general. Although there was a statistically significant correlation between health literacy level and physical functioning which persisted after adjusting for other variables, the strength of this correlation was weak and only had an explanatory power of less than 4%. No significant impact of health literacy was found on the remaining seven SF-36 domains.

However, despite the advantage of convenience in administering REALM, researchers have questioned on the accuracy and sensitivity of using REALM to reflect actual health literacy levels, as REALM is only a word recognition test. Hence, it might fail to measure patients' comprehension capabilities, that is, functional health literacy levels. Besides, there was no functional health literacy test that was readily available to be used among general public, who are also the important target audience in various disease management programs such as disease education and prevention programs. Hence, in chapter 6, we developed and validated a generic functional health literacy test (GFHLT) for use among the English-speaking general public. The 21-item GFHLT demonstrated adequate reliability, good face, content and construct validities and sufficient item discrimination power. Besides, the average completion

was only about 3 minutes, showing promising feature of a screening test.

10.2.3.2 Health psychology-related factors

In the 18-month follow up study of patients undergoing TKR (Chapter 7), the presence of response shift was detected and its impact on both HRQoL and utility assessment was quantified using the then-test approach. We found that response shift was present at both baseline and six months after TKR, and significantly influenced HRQoL scores. It suggested that treatment effect may be masked by response shift, due to gradual adaptation to an improving health status of patients. This may have an impact on the use of conventional pre- and post-test methods to assess improvement in HRQoL and utility scores, which could generate further impact on CUA evaluation of longitudinal disease management programs.

In the application of CUA, patients actually play a very important role as health utilities are usually generated by their subjective assessment of different health states. Hence, it is necessary to understand the variances in health utility values from a psychological perspective. In the exploratory study to investigate the power of expectancy-value model (EVM) in explaining health preferences (Chapter 8), we found that compared with models that only incorporated external variables, EVM that incorporated both external variables and attitudinal attributes could explain a much larger proportion of the variances in health preferences. Besides, EVM could further explore the contributions of each component, either attitudinal attribute or external variable, to identify important factors that influence results of health preferences-based health utilities and quantify their magnitude as well.

Last but not least, we also developed and validated a 12-item scale to measure patients' trust in pharmacists, as such factor could influence quite a number of aspects (such as treatment outcomes, health-related quality of life, health utilizes, treatment cost), which would have a final impact on the CUA results. Based on our results, the 12-item scale was found to be constructed on three factors, namely, technical competence, benevolence and global trust. The scale demonstrated high reliability and good construct validity.

10.3 Main contributions

In general, the results from the studies performed for this thesis would contribute to the improved understanding of both the theoretical and practical issues that affect the feasibility, reliability and robustness of applying CUA analysis in assessing disease management programs in Asian countries. The main contributions of the studies are summarized as follows:

First and foremost, to the best of our knowledge, we provided a pioneering example of using the universalist approach to address how to properly overcome the linguistic and cultural barrier in translating and adapting a measure to ensure its reliability and validity in a different population. Such approach could be further applied by researchers in Asia when they plan to incorporate certain foreign measures for data collection on some parameters for cost-utility analyses.

Second, the factors as identified from the qualitative review served as a comprehensive (albeit not complete) reference list to Asian researchers and decision

makers when conducting CUA. With this reference list, they could then selectively incorporating potentially influential factors in the sensitivity analyses to test the robustness of the results based on different scenarios. This potentially would improve the acceptability of CUA for assessment of cost-effectiveness of disease management programs by decision makers.

Third, in this thesis, we also initiated the exploration of other new factors and the findings contributed to the further understanding and potential expansion of the reference list that we have generated. Although theoretically speaking, factors such as disease knowledge and health literacy may affect patients' empowerment capabilities in disease management, yet their impact on the final CUA results might be quite minimal as such factors are more likely to exert indirect influences. Comparatively, health psychology-related factors tended to play a more influential role. Response shift was suggested as an important factor to be included into the reference list due to its impact on longitudinal assessment that would often be used in CUA of disease management programs. Furthermore, we successfully applied the expectancy-value model to explain variances in health preferences, which could be used as a complementary tool to understand the differences in health utility values that would be generated from various individuals or populations.

Last but not least, we also developed and validated two new scales from scratch, such as GFHLT and patients' trust in pharmacists, which built up a new platform for further exploration of their impact on CUA or other future studies that require the measurement of these two factors.

10.4 Main limitations

The limitations have been discussed in details in the individual chapters. The main limitation we would like to address here is the sampling issue. Due to the constraint of project timeline and limited resources, we were not able to apply probability sampling or purposive sampling to recruit a more representative sample with a substantially larger number of respondents in most of our studies. Hence, it is recommended that our results should be considered as preliminary findings. Cautions should be taken when generalizing our results to a larger and more diversified study population. Besides, due to the lack of knowledge in Malay or Tamil languages and limited financial resources, the study subjects we recruited were either Chinese-speaking or English-speaking Singaporeans. Hence, the findings of the new factors might not be readily applicable to those Malay or Tamil-speaking population in Asia without further verification.

10.5 Future studies

The findings and limitations in the thesis have raised some new concerns and research questions that could be further investigated in future studies.

(1) What are the other factors that will impact the application of CUA in assessing disease management in Asia? Based on the two new scales we have developed in the thesis, future studies could focus on the exploration of the influence of functional health literacy and patients' trust in pharmacists. Besides, any other new factor, not restricted to the two categories (namely, patient-empowerment strategies and health psychology) as we proposed in the thesis, could be investigated to further

verify and expand the current reference list of the influential factors.

(2) What would be the impact of the factors (disease knowledge, health literacy, response shift, and expectancy-value) as we explored on a different population in Asia? Those new results could further verify our findings or lead to the modification of findings in that particular population.

(3) It would also be meaningful and interesting to perform actual CUA studies to assess disease management programs in Asia, in which magnitude of impact of various factors in the reference list would be examined to improve understanding in real case scenarios.

We believe the above studies would further complement the findings in this thesis. They would continue to contribute to the emerging and robust application of CUAs to assess disease management programs in Asia countries for the sustainability of health care resource allocations.

Bibliography

Adili F, Larijani B, Haghighatpanah M. Diabetic patients: Psychological aspects. Ann N Y Acad Sci. 2006; 1084:329-49.

Ahmed S, Mayo NE, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Change in quality of life of people with stroke over time: true change or response shift? Qual Life Res. 2005;14(3):611-27.

Ahmed S, Mayo NE, Wood-Dauphinee S, Hanley JA, Cohen SR. Using the Patient Generated Index to evaluate response shift post-stroke. Qual Life Res. 2005;14(10):2247-57.

Aiken LR. Psychological testing and assessment. Allyn and Bacon , Boston, 1997.

Aldridge MD. Writing and designing readable patient education materials. Nephrol Nurs J. 2004;31(4):373-7.

Allaire SH. Update on work disability in rheumatic diseases. Curr Opin Rheumatol. 2001;13(2):93-8.

Anderson JM. Empowering patients: issues and strategies. Soc Sci Med. 1996; 43(5):697-705.

Anderson LA, Dedrick RF. Development of the trust in physician scale: a measure to assess interpersonal trust in patient-physician relationships. Psychol Rep 1990; 67: 1091-100

Andrus MR, Roth MT. Health literacy: A Review. Pharmacotherapy. 2002; 22(3):282–302.

Anell A, Persson U. Reimbursement and clinical guidance for pharmaceuticals in Sweden: do health-economic evaluations support decision making? Eur J Health Econ. 2005 ;6(3):274-9

Arguedas MR, Chen VK, Eloubeidi MA, Fallon MB. Screening for hepatocellular carcinoma in patients with hepatitis C cirrhosis: a cost-utility analysis. Am J Gastroenterol. 2003;98(3):679-90.

Armstrong TS, Cohen MZ, Eriksen L, Cleeland C. Content validity of self-report measurement instruments: an illustration from the development of the Brain Tumor Module of the M.D. Anderson Symptom Inventory. Oncol Nurs Forum. 2005;32(3):669-76.

Arnesen T, Trommald M. Are QALYs based on time trade-off comparable?--A systematic review of TTO methodologies. Health Econ. 2005;14(1):39-53

Aujoulat I, d'Hoore W, Deccache A. Patient empowerment in theory and practice: polysemy or cacophony? Patient Educ Couns. 2007;66(1):13-20.

Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns. 1999 ;38(1):33-42.

Baker DW. Reading between the lines: deciphering the connections between literacy and health. J Gen Intern Med 1995;14:315–17.

Bakhai A, Stone GW, Grines CL, Murphy SA, Githiora L, Berezin RH, Cox DA, Stuckey T, Griffin JJ, Tcheng JE, Cohen DJ; CADILLAC Investigators. Cost-effectiveness of coronary

stenting and abciximab for patients with acute myocardial infarction: results from the CADILLAC (Controlled Abciximab and Device Investigation to Lower Late Angioplasty Complications) trial. Circulation. 2003;108(23):2857-63.

Baladi JF, Menon D, Otten N. An economic evaluation of finasteride for treatment of benign prostatic hyperplasia. Pharmacoeconomics. 1996;9 (5):443-54

Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.

Belouet C, Bonnichon P, Douard MC, Maroudy D, Vidal-Trecan G. Elements for a cost/utility analysis of long term intravenous devices. Pathol Biol (Paris). 1999;47(3):282-7. French

Beney J, Bero LA, Bond C. Expanding the roles of outpatient pharmacists: effects on health services utilisation, costs, and patient outcomes. Cochrane Database Syst Rev. 2000;(3):CD000336.

Bennet IM. Some questions about screening for literacy. Ann Fam Med. 12 December 2005

Berger ML, Bingefors K, Hedblom EC, Pashos CL, Torrance GW, Dix Smith Meds. Health Care Cost, Quality, and Outcomes: ISPOR Book of Terms. Lawrenceville, NJ: ISPOR, 2003.

Bernhard J, Lowy A, Maibach R, Hürny C; Swiss Group for Clinical Cancer Research (SAKK) and the Swiss Institute for Applied Cancer Research (SIAK), Bern, Switzerland. Response shift in the perception of health for utility evaluation. an explorative investigation. Eur J Cancer. 2001;37(14):1729-35.

Beun JG.. The chronically ill patient and the pharmacist. Pharmaceutical care from the patient's perspective. Pharm World Sci. 1996;18(6):244-6.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1986;i:307-310.

Bonneux L, Birnie E. The discount rate in the economic evaluation of prevention: a thought experiment. J Epidemiol Community Health. 2001;55(2):123-5

Bradley C, Todd C, Gorton T, Symonds E, Martin A, Plowright R. The development of an individualized questionnaire measure of perceived impact of diabetes on quality of life: the ADDQoL. Qual Life Res. 1999; 8(1-2):79-91.

Brander VA, Chang HJ, Chang RW. Rehabilitation and biomechanics. Curr Opin Rheumatol. 1992;4(2):174-8.

Bratzler DW, Nsa W, Houck PM. Performance measures for pneumonia: are they valuable, and are process measures adequate? Curr Opin Infect Dis. 2007 ;20(2):182-9.

Bravata DM, Nelson LM, Garber AM, Goldstein MK. Invariance and inconsistency in utility ratings. Med Decis Making. 2005;25(2):158-67.

Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from

the SF-36. j Health Econ. 2002; 21:271-92

Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ. 2004;13(9):873-84.

Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. J Clin Epidemiol. 1998; 51:1115-128

Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. J Clin Epidemiol. 1998; 51:1115-28.

Breetvelt IS, Van Dam FS. Underreporting by cancer patients: the case of response-shift. Soc Sci Med. 1991;32(9):981-7.

Briggs A, Sculpher M. Sensitivity analysis in economic evaluation: a review of published studies. Health Econ. 1995;4(5):355-71.

Brooks R, with the EuroQol Group: EuroQol: the current state of play. Health Policy 1996; 37:53-72

Brossart DF, Clay DL, Willson VL. Methodological and statistical considerations for threats to internal validity in pediatric outcome data: response shift in self-report outcomes. J Pediatr Psychol. 2002;27(1):97-107.

Brown GC, Brown MM, Sharma S, Busbee B, Brown H. Incremental cost-effectiveness of laser therapy for visual loss secondary to branch retinal vein occlusion. Ophthalmic Epidemiol. 2002;9(1):1-10.

Brown MM, Brown GC, Sharma S, Landy J. Health care economic analyses and value-based medicine. Surv Ophthalmol. 2003;48(2):204-23.

Brown MM. Health care economic analyses. Curr Opin Ophthalmol. 2003;14(3):117-21.

Bruce AT. The handbook of social work research methods. California. Sage Publications, 2001

Brunner HI, Giannini EH. Health-related quality of life in children with rheumatic diseases. Curr Opin Rheumatol. 2003;15(5):602-12.

Buchbinder R, Hall S, Grant G, Mylvaganam A, Patrick MR. Readability and content of supplementary written drug information for patients used by Australian rheumatologists. Med J Aus 2000; 174:575–8.

Buchbinder R, Hall S, Youd JM. Functional health literacy of patients with rheumatoid arthritis attending a community-based rheumatology practice. J Rheumatol. 2006;33(5):879-86

Buxton MJ, Lacey LA, Feagan BG, Niecko T, Miller DW, Townsend RJ. Mapping from disease-specific measures to utility: an analysis of the relationships between the Inflammatory Bowel Disease Questionnaire and Crohn's Disease Activity Index in Crohn's disease and measures of utility. Value Health. 2007;10(3):214-20.

Byrne MM, O'Malley KJ, Suarez-Almazor ME. Ethnic differences in health preferences: analysis using willingness-to-pay. J Rheumatol 2004;31:1811–8.

Cardarelli C, Cereda C, Masiero L, Viscardi E, Faggin R, Laverda A, et al. Evaluation of health status and health-related quality of life in a cohort of Italian children following treatment for a primary brain tumor. Pediatr Blood Cancer 2006; 46:637–44.

Carey M. Diabetes guidelines, outcomes, and cost-effectiveness study: a protocol, prototype, and paradigm. J Am Diet Assoc. 1995; 95(9):976-8

Casey R, Tarride JE, Keresteci MA, Torrance GW. The Erectile Function Visual Analog Scale (EF-VAS): a disease-specific utility instrument for the assessment of erectile function. Can J Urol. 2006;13(2):3016-25

Chan YM, Leung TN, Leung TY, et al. The utility assessment of Chinese pregnant women towards the birth of a baby with Down syndrome compared to a procedure-related miscarriage. Prenat Diagn 2006;26:819–24.

Cheah J. Chronic disease management: a Singapore perspective. BMJ. 2001; 323: (7319): 990-3.

Clark JS. An aging population with chronic disease compels new delivery systems focused on new structures and practices. Nurs Adm Q. 2004; 28(2):105-15.

Clark WF, Churchill DN, Forwell L, Macdonald G, Foster S. To pay or not to pay? A decision and cost-utility analysis of angiotensin-converting-enzyme inhibitor therapy for diabetic nephropathy. CMAJ. 2000;162(2):195-8.

Claudio L, Stingone JA, Godbold J. Prevalence of childhood asthma in urban communities: the impact of ethnicity and income. Ann Epidemiol. 2006;16(5):332-40.

Cohen BJ. Discounting in cost-utility analysis of healthcare interventions: reassessing current practice. Pharmacoeconomics. 2003;21(2):75-87.

Cohen BJ. Statistical power analysis for the behavioral sciences (2nd ed.). 1988. New Jersey: Lawrence Erlbaum.

Cook KF, Ashton CM, Byrne MM, Brody B, Geraci J, Giesler RB, Hanita M, Souchek J, Wray NP. A psychometric analysis of the measurement level of the rating scale, time trade-off, and standard gamble. Soc Sci Med. 2001;53(10):1275-85.

Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. Pharmacoeconomics. 2000;17(1):13-35.

Couch JB. The health care professional's guide to disease management: patient-centered care for the 21st century. Aspen Publishers, Inc. Gaithersburg, Maryland. USA. 1998

Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;22:293–96.

Danese MD, Powe NR, Sawin CT, Ladenson PW. Screening for mild thyroid failure at the periodic health examination: a decision and cost-effectiveness analysis. JAMA. 1996;276(4):285-92.

Davies AR, Ware JE. GHAA's consumer satisfaction survey. Washington, DC: Group Health Association of America, 1991

Davis TC, Crouch MA, Long SW, Jackson RH, Bates P, George RB, Bairnsfather LE. Rapid assessment of literacy levels of adult primary care patients. Fam Med. 1991;23(6):433-5.

Davis TC, Wolf MS. Health literacy: implications for family medicine. Fam Med. 2004;36(8):595-8.

Day JL. Diabetic patient education: determinants of success. Diabetes Metab Res Rev. 2000;16 Suppl 1:S70-4

De Neeling JN. Cost-utility analysis; uncertainties restrict applicability. Ned Tijdschr Geneeskd. 2004;148(22):1106-10. Dutch

Denzin, NK, Lincoln YS. Handbook of qualitative research (2nd edition).Thousand Oaks, CA: Sage Publications. 2000

Department of Statistics. Year Book of Statistics. Singapore. Available from website: http://www.singstat.gov.sg/pubn/reference/yos/statsT-deomography.pdf. Accessed on June 28, 2007.

Dion MJ, Tousignant P, Bourbeau J, Tousignant P, Bourbeau J, Menzies D, Schwartzman K. Measurement of health preferences among patients with tuberculous infection and disease. Med Decis Making 2002; 22 (Suppl. 5):S102–14.

Dolan P, Roberts J. Modeling valuations for EQ-5D health states: an alternative model using differences in valuations. Med Care 2002; 40:442-46

Dolan P. Modeling valuations for EuroQol health states Med Care 1997; 35: 1095-108

Donabedian A. Evaluating the quality of medical care. Milbank Mern. Fund Quart. 44:166-206, 1966

Donabedian A. The definition of quality and approaches to its assessment. Ann Arbor, MI: Health Administration Press; 1980.

Donaldson NE, Rutledge DN, Ashley J. Outcomes of adoption: measuring evidence uptake by individuals and organizations.Worldviews Evid Based Nurs. 2004;1 Suppl 1:S41-51.

Dowse R, Ehlers M. Medicine labels incorporating pictograms: do they influence understanding and adherence? Patient Educ Couns. 2005;58(1):63-70

Dranitsaris G, Hsu T. Cost utility analysis of prophylactic pamidronate for the prevention of skeletal related events in patients with advanced breast cancer. Support Care Cancer. 1999 ;7(4):271-9.

Drasgow F, Kanfer R. Equivalence of psychological measurement in heterogeneous populations. J Appl Psychol. 1985;70:662–680.

Du Toit R, Pritchard N, Heffernan S, Simpson L, Fonn D. A comparison of three different

scales for rating contact lens handling. Optom Vis Sci 2002;79:313–20.

Dugan E, Trachtenberg F, Hall MA. Development of abbreviated measures to assess patient trust in a physician, a health insurer, and the medical profession. BMC Health Serv Res. 2005;5:64.

Eagly AH, Chaiken S. The Psychology of Attitudes. Fortworth, TX: Harcourt Brace Jovanovich College Publishers, 1993.

Earle CC, Kwok A, Gazelle GS, Fuchs CS. Two schedules of second-line irinotecan for metastatic colon carcinoma. Cancer. 2004;101(11):2533-9.

East West Centre, 2002. The Future of Population in Asia: Asia's aging population.

Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, Wentz R, Kwan I, Cooper R. Methods to increase response rates to postal questionnaires. Cochrane Database Syst Rev. 2007 Apr 18;(2):MR000008

El Achhab Y, Nejjari C, Chikri M, Lyoussi B. Disease-specific health-related quality of life instruments among adults diabetic: A systematic review. Diabetes Res Clin Pract. 2008;80(2):171-84

Ethgen O, Bruyère O, Richy F, Dardennes C, Reginster JY. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. J Bone Joint Surg Am. 2004;86-A(5):963-74.

Eurich DT, Johnson JA, Reid KJ, Spertus JA. Assessing responsiveness of generic and specific health related quality of life measures in heart failure. Health Qual Life Outcomes. 2006 24;4:89.

EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199–208.

Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA. 2002;287(20):2691-700.

Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, Denton M, Boyle M. Multiattribute and single-attribute utility functions for the Health Utilities Index mark 3 system. Med Care. 2002; 40(2):113-28.

Fern EF. Advanced Focus Group Approach. Thousand Oaks, CA: Sage, 2001.

Fernandes R. Managing health care costs within an integrated framework. Health Pap. 2002;3(1):70-6.

Fetterolf D, Wennberg D, Devries A. Estimating the return on investment in disease management programs using a pre-post analysis. Dis Manag. 2004;7(1):5-23.

Fishbein M, Ajzen I. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading, MA: Addison-Wesley, 1975.

Fishbein M, Raven BH. The AB scales: an operational definition of belief and attitude. Hum Relat 1962; 15:35–44.

Fitzgerald JT, Funnell MM, Hess GE, Barr PA, Anderson RM, Hiss RG, Davis WK. The reliability and validity of a brief diabetes knowledge test. Diabetes Care. 1998;21(5):706-10.

Fitzner K, Sidorov J, Fetterolf D, Wennberg D, Eisenberg E, Cousins M, et al., Quality and Research Committee, Disease Management Association of America. Principles for assessing disease management outcomes. Dis Manag. 2004; 7(3):191-201.

Fleurence RL, Hollenbeak CS. Rates and probabilities in economic modelling: transformation, translation and appropriate application. Pharmacoeconomics. 2007;25(1):3-6.

Fox-Rushby JA, Hanson K. Calculating and presenting disability adjusted life years (DALYs) in cost-effectiveness analysis. Health Policy Plan.2001; 16:326–31

Franic DM, Pathak DS. Effect of including (versus excluding) fates worse than death on utility measurement. Int J Technol Assess Health Care 2003;19:347– 61.

Friedman DB, Hoffman-Goetz L. A systematic review of readability and comprehension instruments used for print and web-based cancer information. Health Educ Behav. 2006;33(3):352-73.

Gerard K. Cost-utility in practice: a policy maker's guide to the state of the art. Health Policy. 1992;21(3):249-79.

Gilmour P, Hunt RA. Total quality management: integrating quality into design, operations and strategy. Cheshire: Longman, 1995

Glazer WM. Formulary decisions and health economics. J Clin Psychiatry. 1998;59 Suppl 19:23-9.

Gold MR, Muennig P. Measure-dependent variation in burden of disease estimates: implications for policy. Med Care. 2002;40(3):260-6.

Gold MR, Siegel JE, Russell LB. Cost-Effectiveness in Health and Medicine. New York, NY: Oxford University Press; 1996

Golembiewski RT, Billingsley K, Yeager S. Measuring change and persistence in human affairs: Types of change generated by OD designs. J Appl Behav Sci. 1976; 12:133–57

Gong DA, Lee JY, Rozier RG, Pahel BT, Richman JA, Vann WF. Development and testing of the Test of Functional Health Literacy in Dentistry (TOFHLiD). J Public Health Dent. 2007;67(2):105-12

Goold SD, Vijan S. Normative issues in cost effectiveness analysis. J Lab Clin Med. 1998;132(5):376-82

Gordon MM, Hampson R, Capell HA, Madhok R. Illiteracy in rheumatoid arthritis patients as determined by the Rapid Estimate of Adult Literacy in Medicine (REALM) score. Rheumatology (Oxford). 2002;41(7):750-4

Gorsuch RL. Factor analysis, 2nd ed. Hillsdale, NJ: Lawsrence Erlbaum Assoc, 1983

Gryskiewicz KA, Gillespie EL, White CM, Kluger J, Gallagher R, Coleman CI. Impact of Hispanic ethnicity on the likelihood of developing postoperative atria fibrillation. Conn Med. 2006; 70 (2): 77-9

Gudex CM, Hawthorne MR, Butler AG, Duffey PO. Measuring patient benefit from botulinum toxin in the treatment of dystonia. Feasibility of cost-utility analysis. Pharmacoeconomics. 1997;12(6):675-84.

Guerra CE, Dominguez F, Shea JA. Literacy and knowledge, attitudes, and behavior about colorectal cancer screening. J Health Commun. 2005;10(7):651-63.

Guidelines on taking medicines. Electronic version available at http://www.hpb.gov.sg/hpb/ere/ere0307.asp?id=24110, accessed on March 6, 2007.

Gunter MJ. The role of the ECHO model in outcomes research and clinical practice improvement. Am J Manag Care. 1999;5(4 Suppl):S217-24

Guyatt G, King D, Stubbing FD, Goldstein R. Generic and specific measurement of health-related quality of life in a clinical trial of respiratory rehabilitation. J Clin Epidemiol.1999; 52: 187–92.

Hahn EA, Cella D, Dobrez DG, Weiss BD, Du H, Lai JS, Victorson D, Garcia SF. The impact of literacy on health-related quality of life measurement and outcomes in cancer outpatients. Qual Life Res.2007;16(3):495-507.

Hart HE, Redekop WK, Bilo HJ, Meyboom-de Jong B, Berg M. Health related quality of life in patients with type I diabetes mellitus: generic & disease-specific measurement. Indian J Med Res. 2007;125(3):203-16.

Hasson D, Arnetz BB. Validation and findings comparing VAS vs. Likert scales for psychosocial measurements. Int Electron J Health Educ 2005;8:178–92.

Health Economics Core Library Recommendations (2003) by National Information Center on Health Services Research and Health Care Technology, United States. http://www.nlm.nih.gov/nichsr/corelib/hecon.html Accessed on June 15, 2005

Health literacy: report of the Council of Scientific Affairs. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. JAMA 1999;281:552-7.

Hearth-Holmes M, Murphy PW, Davis TC, Nandy I, Elder CG, Broadwell LH, Wolf RE. Literacy in patients with a chronic disease: systemic lupus erythematosus and the reading level of patient education materials. J Rheumatol. 1997;24(12):2335-9

Helper CD, Strand LM. Opportunities and responsibilities in pharmaceutical care. Am J Hosp Pharm. 1990;47(3):533-43.

Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: The universalist approach. Qual Life Res. 1998; 7:323-35

Hermansen CJ, Wilderholt JB. Pharmacist-patient relationship development in an ambulatory clinic setting. Health Commun. 2001;13(3):307-25.

Hinton PR, Brownlow C, McMurray I, Cozens B. SPSS explained. London; New York : Routledge, 2004.

Ho J, Bender BG, Gavin LA, O'Connor SL, Wamboldt MZ, Wamboldt FS. Relations among asthma knowledge, treatment adherence, and outcome. J Allergy Clin Immunol. 2003;111(3):498-502.

Hoey H, Aanstoot HJ, Chiarelli F, Daneman D, Danne T, et al.. Good metabolic control is associated with better quality of life in 2,101 adolescents with type 1 diabetes. Diabetes Care. 2001;24(11):1923-8.

Hoffmann C, Graf von der Schulenburg JM. The influence of economic evaluation studies on decision making. A European survey. The EUROMET group. Health Policy.2000; 52 (3):179-92

Hoffmann C, Stoykova BA, Nixon J, Glanville JM, Misso K, Drummond MF. Do health-care decision makers find economic evaluations useful? The findings of focus group research in UK health authorities. Value Health. 2002;5(2):71-8

Hoffmann T, Worrall L. Designing effective written health education materials: considerations for health professionals. Disabil Rehabil. 2004;26(19):1166-73

Holly E. The focus group research handbook. NTC Business books. Contemporary publishing group. 1999. Lincolnwood Ill.

Homik JE, Suarez-Almazor M. An economic approach to health care. Best Pract Res Clin Rheumatol. 2004;18(2):203-18

Howard DH, Gazmararian J, Parker RM. The impact of low health literacy on the medical costs of Medicare managed care enrollees. Am J Med. 2005;118(4):371-7.

Ikegami N, Drummond M, Fukuhara S, Nishimura S, Torrance GW, Schubert F. Why has the use of health economic evaluation in Japan lagged behind that in other developed countries? Pharmacoeconomics. 2002;20 Suppl 2:1-7.

Inadomi JM. Decision analysis and economic modelling: a primer. Eur J Gastroenterol Hepatol. 2004;16(6):535-42.
International Diabetes Federation. Diabetes Atlas, third edition, 2006.

Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10:407–15.

Johannesson M. Theory and methods of economic evaluation of health care. Kluwer Academic Publishers. 1996

Johnson-Masotti AP, Pinkerton SD, Kelly JA, Stevenson LY. Cost-effectiveness of an HIV risk reduction intervention for adults with severe mental illness. AIDS Care. 2000 ;12(3):321-32.

Johnston SC, Gress DR, Kahn JG. Which unruptured cerebral aneurysms should be treated? A cost-utility analysis. Neurology. 1999;52(9):1806-15

Kanuha VK. The impact of sexuality and race/ethnicity on HIV/AIDS risk among Asian and Pacific Island American (A/PIA) gay and bisexual men in Hawai'i. AIDS Educ Prev. 2000;12(6):505-18.

Kao AC, Green DC, Zaslavsky AM, Koplan JP, Cleary PD. The relationship between method of physician payment and patient trust. JAMA. 1998; 280(19):1708-14.

Kavanaugh A. Economic issues with new rheumatologic therapeutics. Curr Opin Rheumatol. 2007;19(3):272-6.

Kemp J. Is there a best outcome measurement for patients with asthma? Ann Allergy Asthma Immunol. 2006;97(4):432-9.

Kiberd BA, Larson T. Estimating the benefits of solitary pancreas transplantation in nonuremic patients with type 1 diabetes mellitus: a theoretical analysis. Transplantation. 2000;70(7):1121-7

Kim JO, Mueller CW. Factor analysis: statistical methods and practical issues. Berverly Hills: Sage, 1978

Kind P. Measuring quality of life in evaluating clinical interventions: an overview. Ann Med. 2001;33(5):323-7.

Kind P. The EuroQoL instrument: an index of health related quality of life. In: Spilker B, ed. Quality of Life and Pharmacoeconomics in Clinical Trials. Philadelphia, PA: Lippincott-Raven Publishers, 1996.

King DK. Reaction and questions for discussion re: the NVS Health Literacy Assessment. Ann Fam Med. 9 December 2005

King JT Jr, Styn MA, Tsevat J, Roberts MS. “Perfect health” versus “disease free”: the impact of anchor point choice on the measurement of preference values and the calculation of disease-specific disutilities. Med Decis Making 2003;23:212–25.

Kleinbeck C. Reaching positive diabetes outcomes for patients with low literacy. Home Health Nurse. 2005;23(1):16-22

Kleinpeter MA. Health literacy affects peritoneal dialysis performance and outcomes. Adv Perit Dial. 2003;19:115-9.

Ko CY, Maggard M, Livingston EH. Evaluating health utility in patients with melanoma, breast cancer, colon cancer, and lung cancer: a nationwide, population-based assessment. J Surg Res 2003;114: 1–5.

Kobelt G. Health economic issues in rheumatoid arthritis. Scand J Rheumatol. 2006;35(6):415-25.

Koh ET, Seow A, Pong LY, Koh WH, Chan L, Howe HS, Lim TH, Low CK. Cross cultural adaptation and validation of the Chinese Health Assessment Questionnaire for use in

rheumatoid arthritis. J Rheumatol. 1998;25(9):1705-8.

Korolija D, Wood-Dauphinee S, Pointner R. Patient-reported outcomes. How important are they? Surg Endosc. 2007;21(4):503-7.

Krabbe PF, Essink-Bot ML, Bonsel GJ. The comparability and reliability of five health-state valuation methods. Soc Sci Med 1997;45:1641–52.

Kraetschmer N, Sharpe N, Urowitz S, Deber RB. How does trust affect patient preferences for participation in decision-making? Health Expect. 2004; 7 (4): 317-26

Krahn M, Bremner KE, Tomlinson G, Ritvo P, Irvine J, Naglie G. Responsiveness of disease-specific and generic utility instruments in prostate cancer patients. Qual Life Res. 2007;16(3):509-22

Kristiansen IS, Falch JA, Andersen L, Aursnes I. Use of alendronate in osteoporosis--is it cost-effective? Tidsskr Nor Laegeforen. 1997;117(18):2619-22.Norwegian

Kumar PN, Rodriguez-French A, Thompson MA, Tashima KT, Averitt D, et al. A prospective, 96-week study of the impact of Trizivir, Combivir/nelfinavir, and lamivudine/stavudine/nelfinavir on lipids, metabolic parameters and efficacy in antiretroviral-naive patients: effect of sex and ethnicity. HIV Med. 2006;7(2):85-98

Kupersmith J, Holmes-Rovner M, Hogan A, Rovner D, Gardiner J. Cost-effectiveness analysis in heart disease, Part II: Preventive therapies. Prog Cardiovasc Dis. 1995;37(4):243-71

Larson I, Schumacher HR. Comparison of literacy level of patients in a VA arthritis center with the reading level required by education materials. Arthritis Care Res 1992; 5:13–6.

Lau C, Qureshi AK, Scott SG. Association between glycemic control and quality of life in diabetes mellitus. J Postgrad Med. 2004; 50(3): 189-94

Lee KS, Brouwer WB, Lee SI, Koo HW. Introducing economic evaluation as a policy tool in Korea: will decision makers get quality information?: a critical review of published Korean economic evaluations. Pharmacoeconomics. 2005; 23 (7): 709-21

Leisen B, Hyman MR. An improved scale for assessing patients' trust in their physician. Health Mark Q. 2001; 19(1):23-42.

Li N, van Agthoven M, Willemse P, Uyl-de Groot C. A cost--utility analysis comparing second-line chemotherapy schemes in patients with metastatic breast cancer. Anticancer Drugs. 2001;12(6):533-40

Liljas B. How to calculate indirect costs in economic evaluations. Pharmacoeconomics. 1998 ;13:1-7.

Lincoln A, Paasche-Orlow MK, Cheng DM, Lloyd-Travaglini C, Caruso C, Saitz R, Samet JH. Impact of health literacy on depressive symptoms and mental health-related: quality of life among adults with addiction. J Gen Intern Med. 2006;21(8):818-22.

Luo N, Chew LH, Fong KY, Koh DR, Ng SC, Yoon KH, Vasoo S, Li SC, Thumboo J. Do

English and Chinese EQ-5D versions demonstrate measurement equivalence? an exploratory study. Health Qual Life Outcomes. 2003;1(1):7.

Luo N, Chew LH, Fong KY, Koh DR, Ng SC, Yoon KH, Vasoo S, Li SC, Thumboo J. Validity and reliability of the EQ-5D self-report questionnaire in English-speaking Asian patients with rheumatic diseases in Singapore. Qual Life Res. 2003;12(1):87-92.

Luscombe FA. Health-related quality of life measurement in type 2 diabetes. Value Health. 2000;3 Suppl 1:15-28

MacKeigan LD, Larson LN. Development and validation of an instrument to measure patient satisfaction with pharmacy services. Med Care. 1989; 27(5):522-36.

Mancuso CA, Rincon M. Impact of health literacy on longitudinal asthma outcomes. J Gen Intern Med. 2006;21(8):813-7.

Mankiw NG. Principles of economics. 2nd ed. Fort Worth: Harcourt College Publishers, 2001
Manns B, Meltzer D, Taub K, Donaldson C. Illustrating the impact of including future costs in economic evaluations: an application to end-stage renal disease care. Health Econ. 2003;12(11):949-58

Mant J. Process versus outcome indicators in the assessment of quality of health care. Int J Qual Health Care. 2001;13:475-80

Mason JM, Mason AR.The generalisability of pharmacoeconomic studies: issues and challenges ahead. Pharmacoeconomics. 2006;24(10):937-45.

Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. Acad. Management Rev. 1995; 20(3):709-34

Mehrez A, Gafni A. Quality adjusted life years, utility theory, and healthy years equivalents. Med Decis Making. 1989; 9:142–59.

Meltzer D, Egleston B, Stoffel D, Dasbach E. Effect of future costs on cost-effectiveness of medical interventions among young adults: the example of intensive therapy for type 1 diabetes mellitus. Med Care. 2000;38(6):679-85

Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. J Health Econ. 1997;16(1):33-64.

Miller DP, Brownlee CD, McCoy TP, Pignone MP. The effect of health literacy on knowledge and receipt of colorectal cancer screening: a survey study. BMC Fam Pract. 2007 Mar 30;8:16

Ministry of Health, Singapore. National Health Survey Singapore: 1998: p 5–9.

Mishoe SC, MacLean JR. Assessment of health-related quality of life. Respir Care. 2001;46(11):1236-57.

Moorman C, Deshpande R, Zaltman G. Factors affecting trust in marketing relationships. Journal of Marketing. 1993; 57 (1): 81-100

Morimoto T, Fukui T. Utilities measured by rating scale, time trade-off, and standard gamble:

review and reference for health care professionals. J Epidemiol. 2002;12(2):160-78.

Morrow D, Clark D, Tu W, Wu J, Weiner M, Steinley D, Murray MD. Correlates of health literacy in patients with chronic heart failure. Gerontologist. 2006 ;46(5):669-76

Mortensen HB et al. Findings from the Hvidore Study Group on Childhood Diabetes: metabolic control and quality of life. Horm Res. 2002; 57 (Suppl 1):117-20.

Muenning P. Designing and conducting cost-effectiveness analyses in medicine and health care. JOSSEY-BASS, CA. 2002

Murray CJ, Acharya AK. Understanding DALYs (disability-adjusted life years). J Health Econ. 1997;16(6):703-30.

Murray CJ. Quantifying the burden of disease: the technical basis for disability-adjusted life years. Bull World Health Organ. 1994;72(3):429-45.

National Center for Education Statistics. National Adult Literacy Survey. www.nces.ed.gov/pubsearch/pubsinfo.asp?pubid=93275. Accessed on March 6, 2007.
Nuijten MJ, Rutten F. The incorporation of potential confounding variables in Markov models. Pharmacoeconomics. 2003;21(13):941-50

Nuijten MJ. Incorporation of statistical uncertainty in health economic modelling studies using second-order Monte Carlo simulations. Pharmacoeconomics. 2004;22 (12):759-69.

Nuijten MJ. Incorporation of uncertainty in health economic modelling studies. Pharmacoeconomics. 2005;23(8):851-3; author reply 853

Nunnally, J. Psychometric Theory. New York: McGraw-Hill, 1978.

Ofman JJ, Badamgarav E, Henning JM, Knight K, Gano AD Jr, Levan RK, Gur-Arie S, Richards MS, Hasselblad V, Weingarten SR. Does disease management improve clinical and economic outcomes in patients with chronic diseases? A systematic review. Am J Med. 2004; 117 (3): 182-92.

O'Hagan A, McCabe C, Akehurst R, Brennan A, Briggs A, Claxton K, Fenwick E, Fryback D, Sculpher M, Spiegelhalter D, Willan A. Incorporation of uncertainty in health economic modelling studies. Pharmacoeconomics. 2005;23(6):529-36

Oort FJ, Visser MR, Sprangers MA. An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. Qual Life Res. 2005;14(3):599-609.

Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. J Gen Intern Med.1995;10 (10):537-41.

Paschalides C, Wearden AJ, Dickens CM. The associations of anxiety, depression and personal illness representations with glycaemic control and health-related quality of life in patients with type 2 diabetes mellitus. J of Psychosomatic Research, 2004,:57(6): 557-564

PausJenssen AM, Singer PA, Detsky AS. Ontario's formulary committee: how

recommendations are made. Pharmacoeconomics. 2003;21(4):285-94

Petitti DB. Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis. Oxford: Oxford University Press, 1994.

Piette JD, Heisler M, Krein S, Kerr EA. The role of patient-physician trust in moderating medication nonadherence due to cost pressures. Arch Intern Med. 2005 Aug 8-22;165(15):1749-55.

Pilot D.F., 1996. Data Analysis & Statistics for Nursing research. Appleton & Lange, Stamford, CT

Pinkerton SD, Holtgrave DR, Jemmott JB 3rd. Economic evaluation of HIV risk reduction intervention in African-American male adolescents. J Acquir Immune Defic Syndr 2000;25(2):164-72.

Pinkerton SD, Johnson-Masotti AP, Otto-Salaj LL, Stevenson LY, Hoffmann RG. Cost-effectiveness of an HIV prevention intervention for mentally ill adults. Ment Health Serv Res. 2001;3(1):45-55

Polsky D, Keating NL, Weeks JC, Schulman KA. Patient choice of breast cancer treatment: impact on health state preferences. Med Care 2002;40:1068– 79.

Polsky D, Mandelblatt JS, Weeks JC, et al. Economic evaluation of breast cancer treatment: considering the value of patient choice. J Clin Oncol 2003;21:1139– 46.

Postulart D, Adang EM. Response shift and adaptation in chronically ill patients. Med Decis Making. 2000;20(2):186-93.

Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group Ann Med 2001; 33: 337-43

Raisch DW. Understanding quality-adjusted life years and their application to pharmacoeconomic research. Ann Pharmacother. 2000;34(7-8):906-14.

Rall CJ, Munshi AD, Stasior DS. Disease management strategies and the gastroenterologist. Gastroenterol Clin North Am. 1997;26(4):873-94.

Ramos-Remus C, Salcedo-Rocha AL, Prieto-Parra RE, Galvan-Villegas F. How important is patient education? Baillieres Best Pract Res Clin Rheumatol. 2000 ;14(4):689-703.

Razmjou H, Yee A, Ford M, Finkelstein JA. Response shift in outcome assessment in patients undergoing total knee arthroplasty. J Bone Joint Surg Am. 2006;88(12):2590-5.

Razmjou H, Schwartz CE, Yee A, Finkelstein JA. Traditional assessment of health outcome following total knee arthroplasty was confounded by response shift phenomenon. J Clin Epidemiol. 2009 Jan;62(1):91-6.

Reeder CE. Overview of pharmacoeconomics and pharmaceutical outcomes evaluations. Am J Health Syst Pharm. 1995;52(19 Suppl 4):S5-8.

Revivki DA and Kaplan RM. Relationship between psychometric and utility-based approaches to the measurement of health-related quality of life. Quality of Life Research. 1993; 2 (6): 477-87.

Richards DM, Irving MH. Cost-utility analysis of home parenteral nutrition. Br J Surg. 1996;83(9):1226-9.

Richardson G, Manca A. Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. Health Econ. 2004;13(12):1203-10

Ring L, Hofer S, Heuston F, Harris D, O'Boyle CA. Response shift masks the treatment impact on patient reported outcomes (PROs): the example of individual quality of life in edentulous patients. Health Qual Life Outcomes. 2005 7;3:55.

Ritterband DR. Disease management: old wine in new bottles? J Health Manag. 2000;45(4):255-66.

Roos EJ, de Graeff A, van Eijkeren MA, Boon TA, Heintz AP. Quality of life after pelvic exenteration. Gynecol Oncol. 2004;93(3):610-4.

Rubio-Terres C, Aristegui Ruiz I, Medina Redondo F, Izquierdo Ayuso G. Cost-utility analysis of multiple sclerosis treatment with glatiramer acetate or interferon beta in Spain. Farm Hosp. 2003;27(3):159-65. Spanish.

Rudd RE, Rosenfeld L, Gall V. Health literacy and arthritis research and practice. Curr Opin Rheumatol. 2007;19(2):97-100.

Russell LB, Sisk JE. Modeling age differences in cost-effectiveness analysis. A review of the literature. Int J Technol Assess Health Care. 2000;16(4):1158-67.

Rutherford J, Holman R, MacDonald J, Taylor A, Jarrett D, Bigrigg A. Low literacy: a hidden problem in family planning clinics. J Fam Plann Reprod Health Care. 2006; 32(4):235-40.

Rye J. Cloze procedures and the teaching of reading. London : Heinemann Educational, 1982.

Safeer RS, Keenan J. Health literacy: the gap between physicians and patients. Am Fam Physician. 2005 ;72(3):463-8.

Saito I, Inami F, Ikebe T, Moriwaki C, Tsubakimoto A, Yonemasu K, Ozawa H. Impact of diabetes on health-related quality of life in a population study in Japan. Diabetes Res Clin Pract. 2006;73(1):51-7

Salkeld G, Davey P, Arnolda G. A critical review of health-related economic evaluations in Australia: implications for health policy. Health Policy. 1995;31(2):111-25.

Sanders GD, Hlatky MA, Every NR, McDonald KM, Heidenreich PA, Parsons LS, Owens DK. Potential cost-effectiveness of prophylactic use of the implantable cardioverter defibrillator or amiodarone after myocardial infarction. Ann Intern Med. 2001;135(10):870-83.

Sawyer MG, Spurrier N, Kennedy D, Martin J. The relationship between the quality of life of children with asthma and family functioning. J Asthma. 2001; 38: 279-84

Schackman BR, Gold HT, Stone PW, Neumann PJ. How often do sensitivity analyses for economic parameters change cost-utility analysis conclusions? Pharmacoeconomics. 2004;22(5):293-300.

Schillinger D, Barton LR, Karter AJ, Wang F, Adler N. Does literacy mediate the relationship between education and health outcomes? A study of a low-income population with diabetes. Public Health Rep. 2006;121(3):245-54

Schillinger D, Grumbach K, Piette J, Wang F, Osmond D, Daher C, Palacios J, Sullivan GD, Bindman AB. Association of health literacy with diabetes outcomes. JAMA. 2002; 288(4):475-82.

Schleinitz MD, Heidenreich PA. Cost utility analysis of sildenafil compared with papaverine-phentolamine injections. Ann Intern Med. 2005;142 (4):251-9.

Schweickhardt A, Larisch A, Fritzsche K. Differentiation of somatizing patients in primary care: why the effects of treatment are always moderate. J Nerv Ment Dis. 2005;193(12):813-9.

Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, Davies LM, Eastwood A. Generalisability in economic evaluation studies in health care: a review and case studies. Health Technol Assess. 2004 ;8(49):iii-iv, 1-192.

Seidenwurm DJ, McDonnell CH 3rd, Raghavan N, Breslau J. Cost utility analysis of radiographic screening for an orbital foreign body before MR imaging. AJNR Am J Neuroradiol. 2000;21(2):426-33.

Shah VA, Gupta SK, Shah KV, et al. Utility scores measure quality of life in patients with visual morbidity due to diabetic retinopathy or ARMD. Ophthalmic Epidemiol 2004;11:43–51.

Sherbourne CD, Kamberg CJ. Social Functioning: Family and Marital Functioning Measures. In: Stewart AL, Ware JE Jr (Eds). Measuring Functioning and Well-Being: The Medical Outcomes Study Approach. Duke University Press 1992: 183-93.

Siegel JE, Torrance GW, Russell LB, Luce BR, Weinstein MC, Gold MR. Guidelines for pharmacoeconomic studies. Recommendations from the panel on cost effectiveness in health and medicine. Panel on cost Effectiveness in Health and Medicine. Pharmacoeconomics. 1997;11(2):159-68.

Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. JAMA. 1996;276(16):1339-41.

Simon LS. The treatment of rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2004;18(4):507-38

Singapore Census of Population, 2000 advance data release No.3: Literacy and Language. Available from http://www.singstat.gov.sg/papers/c2000/adr-literacy.pdf. Accessed on March 19, 2005

Speros C. Health literacy: concept analysis. J Adv Nurs. 2005;50(6):633-40

Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life

research: a theoretical model. Soc Sci Med. 1999;48(11):1507-15.

Sprangers MA, Schwartz CE. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Soc Sci Med. 1999;48(11):1531-48.

Steele DJ, Blackwell B, Gutmann MC, Jackson TC. The activated patient: dogma, dream, or desideratum? Patient Educ Couns. 1987 ;10(1):3-23.

Steuten L, Vrijhoef B, Severens H, van Merode F, Spreeuwenberg C. Are we measuring what matters in health technology assessment of disease management? Systematic literature review. Int J Technol Assess Health Care. 2006;22(1):47-57.

Stewart DW. The application and misapplication of factor analysis in marketing research. J Mark Res. 1981; 18:51

Stolk EA, Busschbach JJ, Caffa M, Meuleman EJ, Rutten FF. Cost utility analysis of sildenafil compared with papaverine-phentolamine injections. BMJ . 2000; 320 (7243): 1165-8

Sutcliffe N, Clarke AE, Levinton C, Frost C, Gordon C, Isenberg DA. Associates of health status in patients with systemic lupus erythematosus. J Rheumatol. 1999;26(11):2352-6.

Swan JE, Bowers MR, Richardson LD. Customer trust in the salesperson: an integrative review and meta-analysis of the empirical literature. Journal of Business Research. 1999; 44 (2): 93-107

The EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199-208.

Thiel de Bocanegra H, Gany F. Good provider, good patient: changing behaviors to eliminate disparities in health care. Am J Manag Care. 2004;10 Spec No:SP20-8

Thom DH, Campbell B. Patient-physician trust: an exploratory study. J Fam Pract. 1997;44(2):169-76.

Thom DH, Hall MA, Pawlson LG. Measuring patients' trust in physicians when assessing quality of care. Health Aff (Millwood). 2004;23(4):124-32.

Thompson-Fawcett MW, Richard CS, O'Connor BI, Cohen Z, McLeod RS. Quality of life is excellent after a pelvic pouch for colitis-associated neoplasia. Dis Colon Rectum. 2000;43(11):1497-502.

Thumboo J, Fong KY, Chan SP, Leong KH, Feng PH, Thio ST, Boey ML. Validation of the medical outcomes study family and marital functioning measures in SLE patients in Singapore. Lupus 1999; 8:514-20.

Thumboo J, Fong KY, Machin D, Chan SP, Leon KH, Feng PH, Thio ST, Boe ML. A community-based study of scaling assumptions and construct validity of the English (UK) and Chinese (HK) SF-36 in Singapore. Qual Life Res. 2001;10(2):175-88.

Timmerman AA, Anteunis LJ, Meesters CM. Response-shift bias and parent-reported quality of life in children with otitis media. Arch Otolaryngol Head Neck Surg. 2003;129(9):987-91.

Todd WE, Nash D. Disease management: a systematic approach to improve patient outcomes. American Hospital Publishing Inc. Chicago. USA. 1997.

Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. Med Care. 1996 ;34(7):702-22

Torrance GW. Preferences for health outcomes and cost-utility analysis. Am J Manag Care. 1997;3 Suppl:S8-20

United Nations (2001) World Population Ageing: 1950–2050. New York, NY, USA: United Nations

Van den Arend IJ, Stolk RP, Rutten GE, Schrijvers GJ. Education integrated into structured general practice care for Type 2 diabetic patients results in sustained improvement of disease knowledge and self-care. Diabet Med. 2000;17(3):190-7.

Van DM, Koot HM, Saad HH, et al. Observational visual analog scale in pediatric pain assessment: useful tool or good riddance? Clin J Pain 2002;18:310– 6.

Van Hout BA, Goes ES, Grijseels EW, van Ufford MA. Economic evaluation in the field of cardiology: theory and practice. Prog Cardiovasc Dis. 1999;42(2):167-73.

Vanelli M, Chiarelli F, Chiari G, Tumini S. Relationship between metabolic control and quality of life in adolescents with type 1 diabetes. Report from two Italian centres for the management of diabetes in childhood. Acta Biomed Ateneo Parmense. 2003; 74 (Suppl 1):13-7.

Villaire M, Mayer G. Low health literacy: the impact on chronic illness management. Prof Case Manag. 2007;12(4):213-6

Walker D, Fox-Rushby J. Allowing for uncertainty in economic evaluations: qualitative sensitivity analysis. Health Policy Plan. 2001;16(4):435-43.

Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qaul Life Res. 2005;14(6): 1523-32.

Ward MM. Outcome measurement: health status and quality of life. Curr Opin Rheumatol. 2004;16(2):96-101

Ware JE. Effects of acquiescent response set on patient satisfaction ratings. Med Care 1978; 16:327

Wee HL, Cheung YB, Fong KY, Luo N, Machin D, Thumboo J Are English- and Chinese-Language versions of the SF-6D equivalent? A comparison from a population-based study. Clin Ther. 2004; 26(7): 1137-148

Wee HL, Ho HK, Li SC. Public awareness of diabetes mellitus in Singapore. Singapore Med J. 2002; 43(3):128-34.

Wee HL, Li SC, Xie F, Zhang XH, Luo N, Feeny D, Cheung YB, Machin D, Fong KY,

Thumboo J. Validity, feasibility and acceptability of time trade-off and standard gamble assessments in health valuation studies: a study in a multiethnic Asian population in Singapore. Value Health. 2008 Mar;11 Suppl 1:S3-10.

Wee HL, Loke WC, Li SC, Fong KY, Cheung YB, Machin D, Luo N, Thumboo J. Cross-cultural adaptation and validation of Singapore Malay and Tamil versions of the EQ-5D. Ann Acad Med Singapore. 2007;36(6):403-8.

Wee HL, Tan CE, Goh SY, Li SC. Usefulness of the Audit of Diabetes-Dependent Quality-of-Life (ADDQoL) questionnaire in patients with diabetes in a multi-ethnic Asian country. Pharmacoeconomics. 2006;24(7):673-82

Wee HL, Zhang XH, Tan K, Thumboo J, Li SC. Reliability and validity of the general diabetes knowledge test. Proceedings of 10th ISPOR Annual Meeting, Washington, DC, USA, May 15-18, 2005.

Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, Mockbee J, Hale FA. Quick assessment of literacy in primary care: the newest vital sign. Ann Fam Med.2005;3(6):514-22

Weiss BD. Health literacy: an important issue for communicating health information to patients. Zhonghua Yi Xue Za Zhi (Taipei). 2001;64(11):603-8.
Wildi SM, Cox MH, Clark LL, et al. Assessment of health state utilities and quality of life in patients with malignant esophageal dysphagia. Am J Gastroenterol 2004;99:1044–9.

Wilson IB. Clinical understanding and clinical implications of response shift. Soc Sci Med. 1999;48(11):1577-88.

Winkelmayer WC, Benner JS, Glynn RJ, et al. Assessing health state utilities in elderly patients at cardiovascular risk. Med Decis Making 2006;26:247–54.

Wolf MS, Gazmararian JA, Baker DW. Health literacy and functional health status among older adults. Arch Intern Med. 2005;165(17):1946-52

Wolfe F. Psychological distress and rheumatic disease. Scand J Rheumatol. 1999;28(3):131-6.

World Health Organization. Diabetes Fact Sheet. 2006. Available from http://www.who.int/mediacentre/factsheets/fs312/en/. Accessed on April 5, 2008

World Health Organization. NHA ratios and per capita levels (Excel). Available from : http://www.who.int/nha/country/en/index.html. Accessed on January 27, 2008.
Wright V. Introduction to rheumatic disease. J R Soc Health. 1990;110(5):153-5.

Xie F, Li SC, Lo NN, Yeo SJ, Yang KY, Yeo W, et al. Cross-cultural adaptation and validation of Singapore English and Chinese Versions of the Oxford Knee Score (OKS) in knee osteoarthritis patients undergoing total knee replacement. Osteoarthritis Cartilage. 2007;15(9):1019-24.

Zwiebel A. Changing educational counsellors' attitudes toward mental retardation: comparison of two measurement techniques. Int J Rehabil Res. 1987;10(4):383-9.

Appendices

Appendix 2.1 The ADDQoL-English version

This questionnaire asks about your *quality of life* – in other words how good or bad you feel your life to be.

Please put a "✓" in the box that best indicates your response for each item.

There are no right or wrong answers. We just want to know how you feel.

1 In general, my present quality of life is:						
❑	❑	❑	❑	❑	❑	❑
excellent	very good	good	neither good nor bad	bad	very bad	extremely bad

Now we would like to know how your quality of life is affected by your diabetes, its management and any complications you may have.

2 If I did not have diabetes, my quality of life would be:				
❑	❑	❑	❑	❑
very much better	much better	a little better	the same	worse

Please respond to the following more specific questions. For each aspect of life described:

For Part (a): put an "X" in one box to show how diabetes affects this aspect of your life;

For Part (b): put an "X" in one box to show how important this aspect of your life is to your quality of life.

3	***Do you have family / relatives?*** yes ❑ If 'yes', please continue with parts (a) and (b) no ❑ If 'no', please go to question 10.				
(a)	**If I did not have diabetes, my family life would be**				
	❑	❑	❑	❑	❑
	very much better	much better	a little better	the same	worse
(b)	**My family life is:**				
	❑	❑	❑	❑	
	very important	important	somewhat important	not at all important	

4	**(a)**	**If I did not have diabetes, my friendships and social life would be:** ❑ very much better ❑ much better ❑ a little better ❑ the same ❑ worse
	(b)	**My friendships and social life are:** ❑ very important ❑ important ❑ somewhat important ❑ not at all important

5		***Do you have or would you like to have a close personal relationship?*** yes ❑ If 'yes', please continue with parts (a) and (b) no ❑ If 'no', please go to question 12.
	(a)	**If I did not have diabetes, my closest personal relationship would be:** ❑ very much better ❑ much better ❑ a little better ❑ the same ❑ worse
	(b)	**For me, having a close personal relationship is:** ❑ very important ❑ important ❑ somewhat important ❑ not at all important

6		***Do you have or would you like a sex life?*** yes ❑ If 'yes', please continue with parts (a) and (b) no ❑ If 'no', please go to question 13.
	(a)	**If I did not have diabetes, my sex life would be:** ❑ very much better ❑ much better ❑ a little better ❑ the same ❑ worse
	(b)	**For me, a sex life is:** ❑ very important ❑ important ❑ somewhat important ❑ not at all important

7	**(a)**	**If I did not have diabetes, my physical appearance would be:** ❑ very much better ❑ much better ❑ a little better ❑ the same ❑ worse
	(b)	**My physical appearance is:** ❑ very important ❑ important ❑ somewhat important ❑ not at all important

8 (a)	**If I did not have diabetes, physically I could do:** ❑ very much more ❑ much more ❑ a little more ❑ the same ❑ less
(b)	**For me, how much I can do physically is:** ❑ very important ❑ important ❑ somewhat important ❑ not at all important

9	***Are you currently working, looking for work or would you like to work?*** yes ❑ If 'yes', please continue with parts (a) and (b) no ❑ If 'no', please go to question 16.
(a)	**If I did not have diabetes, my working life would be:** ❑ very much better ❑ much better ❑ a little better ❑ the same ❑ worse
(b)	**For me, working life is:** ❑ very important ❑ important ❑ somewhat important ❑ not at all important

10	***Do you ever go on holiday or want to go on holiday?*** yes ❑ If 'yes', please continue with parts (a) and (b) no ❑ If 'no', please go to question 17.
(a)	**If I did not have diabetes, my holidays would be:** ❑ very much better ❑ much better ❑ a little better ❑ the same ❑ worse
(b)	**For me, holidays are:** ❑ very important ❑ important ❑ somewhat important ❑ not at all important

11 (a)	**If I did not have diabetes, I would enjoy my leisure activities:** ❑ very much more ❑ much more ❑ a little more ❑ the same ❑ less
(b)	**My leisure activities are:** ❑ very important ❑ important ❑ somewhat important ❑ not at all important

12 (a)	**If I did not have diabetes, local or long distance journeys would be:** ❑ very much easier ❑ much easier ❑ a little easier ❑ the same ❑ more difficult
(b)	**For me, local or long distance journeys are:** ❑ very important ❑ important ❑ somewhat important ❑ not at all important

13 (a)	**If I did not have diabetes, my self-confidence would be:** ❑ very much greater ❑ much greater ❑ a little greater ❑ the same ❑ less
(b)	**My self-confidence is:** ❑ very important ❑ important ❑ somewhat important ❑ not at all important

14 (a)	**If I did not have diabetes, my motivation would be:** ❑ very much greater ❑ much greater ❑ a little greater ❑ the same ❑ less
(b)	**My motivation is:** ❑ very important ❑ important ❑ somewhat important ❑ not at all important

15 (a)	**If I did not have diabetes, the way people in general react to me would be:** ❑ very much better ❑ much better ❑ a little better ❑ the same ❑ worse
(b)	**The way people in general react to me is:** ❑ very important ❑ important ❑ somewhat important ❑ not at all important
16 (a)	**If I did not have diabetes, my feelings about the future (e.g. worries, hopes) would be:** ❑ very much better ❑ much better ❑ a little better ❑ the same ❑ worse
(b)	**My feelings about the future are:** ❑ very important ❑ important ❑ somewhat important ❑ not at all important

17	**(a)**	**If I did not have diabetes, my financial situation would be:**				
		❑ very much better	❑ much better	❑ a little better	❑ the same	❑ worse
	(b)	**My financial situation is:**				
		❑ very important	❑ important	❑ somewhat important	❑ not at all important	

18	**(a)**	**If I did not have diabetes, I would have to depend on others (when I do not want to):**				
		❑ very much less	❑ much less	❑ a little less	❑ the same	❑ more
	(b)	**For me, not having to depend on others is:**				
		❑ very important	❑ important	❑ somewhat important	❑ not at all important	

19	**(a)**	**If I did not have diabetes, my living conditions would be:**				
		❑ very much better	❑ much better	❑ a little better	❑ the same	❑ worse
	(b)	**My living conditions are:**				
		❑ very important	❑ important	❑ somewhat important	❑ not at all important	

20	**(a)**	**If I did not have diabetes, my freedom to eat as I wish would be:**				
		❑ very much greater	❑ much greater	❑ a little greater	❑ the same	❑ less
	(b)	**My freedom to eat as I wish is:**				
		❑ very important	❑ important	❑ somewhat important	❑ not at all important	
21	**(a)**	**If I did not have diabetes, my freedom to drink as I wish (e.g. sweetened hot and cold drinks, fruit juice, alcohol) would be:**				
		❑ very much greater	❑ much greater	❑ a little greater	❑ the same	❑ less
	(b)	**My freedom to drink as I wish is:**				
		❑ very important	❑ important	❑ somewhat important	❑ not at all important	

Appendix 2.2 The ADDQoL-Chinese version

这个问卷是有关您的生活素质。换句话说，就是你觉得你的生活有多好或多差。
对每个问题，请在最适合你的答案的那个方格内打一个“✓”。
这些问题没有对或错的答案。我们想知道的是你觉得现在的生活如何。

1 总得来说， 我现在的生活素质是:

☐	☐	☐	☐	☐	☐	☐
特别好	很好	好	不好也不差	差	很差	特别差

现在我们想知道你的生活素质是如何被你的糖尿病，有关护理（包括服用药物、看医生、和任何饮食方面的调整）以及任何并发症所影响。

2 如果我没有糖尿病, 我的生活素质将会:

☐	☐	☐	☐	☐
好特别多	好很多	好一点	一样	更差

请回答以下更详细的问题。对于被描述的生活的每个方面:

对问题(a)： 请在一个方格内打一个“X”，来表示糖尿病如何影响你这方面的生活；

对问题(b)：请在一个方格内打一个“X”，来表示这方面的生活对你的生活素质有多重要。

3	*你有家人或亲属吗?* 有 ☐ 如果有,请接着回答(a)和(b) 没有 ☐ 如果没有,请前往第 10 题
(a)	如果我没有糖尿病, 我的家庭生活将会： ☐ 好特别多 ☐ 好很多 ☐ 好一点 ☐ 一样 ☐ 更差
(b)	我的家庭生活： ☐ 很重要 ☐ 重要 ☐ 有些重要 ☐ 根本不重要

4	**(a)**	如果我没有糖尿病, 我的友谊和社交生活将会： ☐ 好特别多　☐ 好很多　☐ 好一点　☐ 一样　☐ 更差
	(b)	我的友谊和社交生活： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

5		*你有没有或想不想有一个关系亲近的人？* 有/ 想 ☐ 如果有,请接着回答(a)和(b) 没有/ 不想 ☐ 如果没有,请前往第 12 题
	(a)	如果我没有糖尿病, 我和我关系最亲近的人将会： ☐ 好特别多　☐ 好很多　☐ 好一点　☐ 一样　☐ 更差
	(b)	对我来说, 有一个关系亲近的人： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

6		*你有没有或想不想有性生活？* 有/ 想 ☐ 如果有,请接着回答(a)和(b) 没有/ 不想 ☐ 如果没有,请前往第 13 题
	(a)	如果我没有糖尿病, 我的性生活将会： ☐ 好特别多　☐ 好很多　☐ 好一点　☐ 一样　☐ 更差
	(b)	对我来说, 拥有性生活： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

7	**(a)**	如果我没有糖尿病,我的外貌将会： ☐ 好特别多　☐ 好很多　☐ 好一点　☐ 一样　☐ 更差
	(b)	我的外貌： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

8	**(a)**	如果我没有糖尿病,我体能上能做的将会： ☐ 增加特别多　☐ 增加很多　☐ 增加一点　☐ 一样　☐ 更少
	(b)	对我来说,能做多少体能活动： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

9	*你目前是否在工作、找工作或想去工作?* 是 ☐ 如果是,请接着回答(a)和(b) 否 ☐ 如果否,请前往第 8 题
(a)	如果我没有糖尿病, 工作生活将会： ☐ 好特别多　☐ 好很多　☐ 好一点　☐ 一样　☐ 更差
(b)	对我来说, 拥有工作生活： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

10	*你曾经去度假或想去度假吗?* 有 ☐ 如果有,请接着回答(a)和(b) 没有 ☐ 如果没有,请前往第 17 题
(a)	如果我没有糖尿病, 我的假期将会： ☐ 好特别多　☐ 好很多　☐ 好一点　☐ 一样　☐ 更差
(b)	对我来说,假期： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

11	**(a)**	如果我没有糖尿病, 我享受消闲活动的乐趣将会： ☐ 增加特别多　☐ 增加很多　☐ 增加一点　☐ 一样　☐ 更少
	(b)	我的消闲活动： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

12 (a)	如果我没有糖尿病, 本地或长途旅程将会： ☐ 特别容易多　☐ 容易多　☐ 容易一点　☐ 一样　☐ 更难
(b)	对我来说, 本地或长途旅程： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

13 (a)	如果我没有糖尿病, 我的自信心将会： ☐ 特别高　☐ 很高　☐ 高一点　☐ 一样　☐ 更少
(b)	我的自信心： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

14 (a)	如果我没有糖尿病, 我的推动力将会： ☐ 特别多　☐ 很多　☐ 多一点　☐ 一样　☐ 更少
(b)	我的推动力： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

15 (a)	如果我没有糖尿病, 别人通常对我的反应将会： ☐ 好特别多　☐ 好很多　☐ 好一点　☐ 一样　☐ 更差
(b)	别人通常对我的反应： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

16 (a)	如果我没有糖尿病, 我对将来的心情 (比如，担心、希望等等) 将会： ☐ 好特别多　☐ 好很多　☐ 好一点　☐ 一样　☐ 更差
(b)	我对将来的心情： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

17 (a)	如果我没有糖尿病, 我的经济情况将会： ☐ 好特别多　☐ 好很多　☐ 好一点　☐ 一样　☐ 更差
(b)	我的经济情况： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

18 (a)	如果我没有糖尿病, 我必须依赖别人(当我不想依赖别人)时将会： ☐ 少特别多　☐ 少很多　☐ 少一点　☐ 一样　☐ 更多
(b)	对我来说, 不须依赖别人： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

19 (a)	如果我没有糖尿病, 我的居住状况将会： ☐ 好特别多　☐ 好很多　☐ 好一点　☐ 一样　☐ 更差
(b)	我的居住状况： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

20 (a)	如果我没有糖尿病, 我想吃什么就吃什么的自由将会： ☐ 增加特别多　☐ 增加很多　☐ 增加一点　☐ 一样　☐ 更少
(b)	我想吃什么就吃什么的自由： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

21 (a)	如果我没有糖尿病, 我想喝什么就喝什么的自由(比如果汁、酒、冷热甜饮料)将会： ☐ 增加特别多　☐ 增加很多　☐ 增加一点　☐ 一样　☐ 更少
(b)	我想喝什么就喝什么的自由(比如果汁、酒、冷热甜饮料)： ☐ 很重要　☐ 重要　☐ 有些重要　☐ 根本不重要

Appendix 4.1
The General Diabetes Knowledge Test

This questionnaire asks about your knowledge of diabetes. Please tick the most appropriate answer.

Question	Yes	No	Unsure
1. General Knowledge of diabetes			
a. Diabetes is a condition of high blood sugar	❑	❑	❑
b. Type 1 diabetes is a condition of insufficient insulin	❑	❑	❑
c. Type 2 diabetes is a condition of the body not responding to insulin	❑	❑	❑
d. Diabetes is non-contagious	❑	❑	❑
e. Diabetes is not curable.	❑	❑	❑
2. Risk factors of diabetes.			
The following are risk factors of diabetes:	❑	❑	❑
a. Family history of diabetes	❑	❑	❑
b. Age above 40 years old	❑	❑	❑
c. Obesity	❑	❑	❑
3. Symptoms of diabetes			
The following are symptoms of diabetes:			
a. Constant feeling of thirst	❑	❑	❑
b. Frequent urination	❑	❑	❑
c. Weight loss despite normal appetite	❑	❑	❑
d. Blurred vision	❑	❑	❑
e. Slow healing of cuts and wounds	❑	❑	❑
f. Tiredness and weakness	❑	❑	❑
4. Complications of diabetes			
The following are complications of diabetes:			
a. Decaying limbs that require surgical removal	❑	❑	❑
b. Eye problems	❑	❑	❑
c. Kidney problems	❑	❑	❑
d. High blood pressure	❑	❑	❑
e. Loss of sensation in arms and legs	❑	❑	❑

Question	Yes	No	Unsure
5. Treatment and management of diabetes.			
a. Insulin injections are available for the control of diabetes	❑	❑	❑
b. Tablets and capsules are available for the control of diabetes	❑	❑	❑
c. Diabetics should carry sweets and jelly beans when they are out	❑	❑	❑
d. Diabetics should exercise regularly	❑	❑	❑
e. Diabetics should have good weight control	❑	❑	❑
f. Diabetics should go for regular eye check-up	❑	❑	❑
g. Diabetics should have a low fat and high fiber diet	❑	❑	❑
h. Diabetics should care for their toes and feet	❑	❑	❑
i. Diabetics should not consume alcohol	❑	❑	❑
j. Diabetics should not donate blood	❑	❑	❑
k. Diabetics should not smoke	❑	❑	❑
l. Diabetics should not wear tight shoe	❑	❑	❑
m. Diabetics should not skip meal when busy	❑	❑	❑
6. Monitoring of diabetes			
a. Diabetics should test for blood glucose	❑	❑	❑
b. Diabetics should test for sugar in the urine	❑	❑	❑
c. Diabetics should make regular visits to the eye doctor	❑	❑	❑
d. Diabetics should go for regular medical check-ups	❑	❑	❑

Appendix 5.1
Rapid Estimate of Adult Literacy in Medicine (REALM)

REALM score sheet

LIST 1	**LIST 2**	**LIST 3**
1) Fat _____	1) Fatigue _____	1) Allergic _____
2) Flu _____	2) Pelvic _____	2) Menstrual _____
3) Pill _____	3) Jaundice _____	3) Testicle _____
4) Dose _____	4) Infection _____	4) Colitis _____
5) Eye _____	5) Exercise _____	5) Emergency _____
6) Stress_____	6) Behavior _____	6) Medication _____
7) Smear _____	7) Prescription _____	7) Occupation _____
8) Nerves _____	8) Notify _____	8) Sexually _____
9) Germs _____	9) Gallbladder _____	9) Alcoholism _____
10)Meals _____	10) Calories _____	10) Irritation _____
11) Disease _____	11) Depression _____	11) Constipation____
12) Cancer_____	12) Miscarriage ____	12) Gonorrhea _____
13) Caffeine_____	13) Pregnancy _____	13) Inflammatory___
14) Attack _____	14) Arthritis _____	14) Diabetes_____
15) Kidney _____	15) Nutrition _____	15) Hepatitis_____
16) Hormones _____	16) Menopause ____	16) Antibiotics _____
17) Herpes _____	17) Appendix _____	17) Diagnosis _____
18) Seizure _____	18) Abnormal _____	18) Potassium _____
19) Bowel _____	19) Syphilis _____	19) Anemia _____
20) Asthma _____	20) Hemorrhoids ___	20) Obesity _____
21) Rectal _____	21) Nausea _____	21) Osteoporosis ___
22) Incest _____	22) Directed _____	22) Impetigo _____
# of (+) in list 1:_____	# of (+) in list 2:_____	# of (+) in list 3:_____

Symbols for scoring: (+) = correct; (-) = word not attempted; (/) = mispronounced word

Appendix 6.1
Generic Functional Health Literacy Test (GFHLT)
Instructions for respondents and research assistants:
1) Please carefully record down your start time and end time! (Research assistant, please help respondents fill in the time if necessary). Please note that maximum completion time is 3 minutes! After 3 minutes, please hand over your answers to the research assistants immediately!
2) The following passage is selected from a health education material published by the Singapore Health Promotion Board. Please circle only 1 word you consider as the most appropriate from the 3 candidate items.

Guidelines on Taking Medicines

1. When you receive medicine from your pharmacist or doctor, you must understand fully how to use your medicine before you leave. Listen carefully to the **(speech / instructions / then)** and check against the **(label / poster / stand)** on the medicine container. **(Wonder/ Ask / Listen)** when you do not **(read/ understand / play)** the instructions or when **(they/ we / yours)** are not clear. You **(need/ should / well)** know the following:
   Name **(thus/ and / or)** strength of the medicine,
   (Aim/ Purpose / Right) of the medicine,
   Directions **(to / for / and)** use,
   Any activity, food and **(others/ other / which)** medicines to be avoided,
   (How/ Any / An) side effects to look **(up/ out/ to)** for.
2. Be sure you **(track/ follow / ignore)** the instructions. Do not **(overwhelm/ exceed / reach)** the stated dose or **(mass/ treatment/ illness)** period.
3. If you are **(eating/ prescribed/ stating)** antibiotics, make sure you **(complete/ end / stop)** the prescription.
4. Discard expired **(medicines/ utensils/ drinks)** as they may no **(shorter/ less/ longer)** be effective.
5. Please always **(interact/ check / ask)** with your pharmacist before **(finishing/ taking / learning)** any over-the-counter medicines.

Instructions of scoring Generic health literacy test (the next page)
1. Please calculate the time of completion: _____ minutes
2. Please calculate the number of correct items _____.

Appendix 8.1
A complete set of questions and answers used to study both "expectancy" and "value" of health state "11121" as an example.

Instruction: Please answer the following questions by circling ONLY ONE number.
1. Please read card 3 carefully.

> ◇ **No problems in walking about**
>
> ◇ **No problems in washing or dressing yourself**
>
> ◇ **No problems in performing usual activities**
>
> ◆ **Moderate pain or discomfort**
>
> ◇ **Not anxious or depressed**

1a) Do you think that living in this health state for the rest of your life will worsen your quality of life in terms of health?

Extremely unlikely	Very unlikely	Somewhat unlikely	Neutral	Somewhat likely	Very likely	Extremely likely
-3	-2	-1	0	1	2	3

1b) You think that worsening of your quality of life in terms of health is _____.

Extremely bad	Very bad	Somewhat bad	Neither good nor bad	Somewhat good	Very good	Extremely good
-3	-2	-1	0	1	2	3

1c) Do you think that living in this health state for the rest of your life will add a burden to your family?

Extremely unlikely	Very unlikely	Somewhat unlikely	Neutral	Somewhat likely	Very likely	Extremely likely
-3	-2	-1	0	1	2	3

1d) You think that adding a burden to your family is _____.

Extremely bad	Very bad	Somewhat bad	Neither good nor bad	Somewhat good	Very good	Extremely good
-3	-2	-1	0	1	2	3

1e) Do you think that living in this health state for the rest of your life will make you less independent?

Extremely unlikely	Very unlikely	Somewhat unlikely	Neutral	Somewhat likely	Very likely	Extremely likely
-3	-2	-1	0	1	2	3

1f) You think that making you less independent is _____.

Extremely bad	Very bad	Somewhat bad	Neither good nor bad	Somewhat good	Very good	Extremely good
-3	-2	-1	0	1	2	3

1g) Do you think that living in this health state for the rest of your life will make you not able to work or study?

Extremely unlikely	Very unlikely	Somewhat unlikely	Neutral	Somewhat likely	Very likely	Extremely likely
-3	-2	-1	0	1	2	3

1h) You think that making you not being able to work or study is _____.

Extremely bad	Very bad	Somewhat bad	Neither good nor bad	Somewhat good	Very good	Extremely good
-3	-2	-1	0	1	2	3

Appendix 9.1

The 18-item scale to measure patients' trust in pharmacists

1. I trust the pharmacist who has updated knowledge.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

2. I trust the pharmacist if he/she evaluates my medical problem thoroughly.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

3. I trust the pharmacist who could keep the information we discuss totally private.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

4. I trust the pharmacist if he/she tells me about a mistake he/she has made on my medication.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

5. I trust the pharmacist if he/she expresses concern and talks to me with reassuring and comforting words.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

6. I trust the pharmacist if he/she could communicate with me clearly and completely.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

7. I trust a pharmacist who shows sufficient respect for me.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

8. I trust the pharmacist if he/she provides me with effective medication at a reasonable price to me.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

9. I trust the pharmacist if he/she allows me to make decision on which medication to take when there are alternatives.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

10. I trust an older pharmacist more than a younger one.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

11. I trust a hospital pharmacist more than other types of pharmacist (community pharmacist, polyclinic pharmacist, etc.).

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

12. I trust a pharmacist with professional appearance.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

13. I will not trust other pharmacists again if I have an unpleasant past experience with a certain pharmacist.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

14. I trust a pharmacist of the same race as myself more than other races.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

15. I trust a pharmacist of the same gender as myself more than the opposite gender.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

16. I trust a pharmacist who has been recommended by my friends, neighbors or relatives.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

17. I trust a pharmacist whose set-up of the counter is neat and tidy.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3

18. I trust a pharmacist so much that I always try to follow his/her advice.

Totally disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Totally agree
-3	-2	-1	0	1	2	3